GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 14, 2002, 16:46:22 ; Search time 35 Seconds
(without alignments)
449.245 Million cell updates/sec

Title: US-09-598-443-2_COPY_1_118
Perfect score: 626
Sequence: 1 MPGVCDRAPDFLSPSEDQVL.....SIQNISFSSFTLQRAGPTSH 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	626	100.0	410	20	AAY25426	Human SIGIRR prote
2	626	100.0	410	22	AAB61139	Human NOV9 protein
3	626	100.0	504	23	AAU86151	Human PRO342 polyp
4	240	38.3	407	22	AAU17408	Novel signal trans
5	240	38.3	407	22	AAM43652	Human polypeptide
6	240	38.3	410	22	AAM43583	Human polypeptide
7	234	37.4	363	21	AAB58416	Lung cancer associ
8	105	16.8	871	22	AAB30570	A splice variant o
9	105	16.8	871	23	AAE16274	Human kinase PKIN-
10	105	16.8	1351	22	AAB30567	Amino acid sequenc
11	105	16.8	1610	22	AAB30568	A full length huma
12	105	16.8	1618	22	AAB85504	Human protein kina
13	105	16.8	1665	23	AAE24151	Human kinase (PKIN
14	105	16.8	1665	23	AAO15372	Human myosin light
15	105	16.8	2596	22	AAB30569	A splice variant o
16	98	15.7	548	22	AAB30572	A full length mous
17	97	15.5	548	22	AAB30571	A full length rat
18	96	15.3	534	22	ABB66941	Drosophila melanog
19	94.5	15.1	295	22	ABB60698	Drosophila melanog
20	90.5	14.5	506	22	ABG10463	Novel human diagno
21	90.5	14.5	524	22	ABB10423	Human cDNA SEQ ID
22	90.5	14.5	524	22	AAU18057	Human immunoglobul
23	90.5	14.5	654	22	ABB10186	Human cDNA SEQ ID
24	90.5	14.5	1694	22	AAE09449	Human sbg24878SIa
25	90.5	14.5	1700	23	ABB05044	Human NOV6 protein
26	90.5	14.5	1709	22	AAE09448	Human sbg24878SIa
27	90.5	14.5	1839	22	ABG10466	Novel human diagno
28	88	14.1	1336	22	ABB57809	Drosophila melanog
29	88	14.1	5635	23	ABP60991	Novel human protei
30	87.5	14.0	3117	21	AAY53667	Sequence gi/332818
31	86	13.7	898	22	ABG12152	Novel human diagno
32	84.5	13.5	739	22	AAB62398	Human MBSP2 polype
33	84	13.4	292	22	ABB10296	Human cDNA SEQ ID
34	84	13.4	315	22	AAU17983	Human immunoglobul
35	84	13.4	465	19	AAW55045	Neural adhesion mo
36	84	13.4	937	22	AAM78714	Human protein SEQ
37	84	13.4	952	22	AAM78715	Human protein SEQ
38	82	13.1	394	21	AAB43943	Human cancer assoc
39	82	13.1	394	23	ABP41879	Human ovarian anti
40	82	13.1	628	22	ABB50259	Lutheran blood gro
41	82	13.1	628	22	ABB50290	Lutheran blood gro
42	81.5	13.0	496	22	ABB62052	Drosophila melanog
43	81	12.9	816	14	AAR39647	Human fibroblast g
44	80	12.8	1571	19	AAW42087	Human Down syndrom
45	80	12.8	1910	19	AAW42086	Human Down syndrom

ALIGNMENTS

RESULT 1
AAY25426
ID AAY25426 standard; Protein; 410 AA.
XX
AC AAY25426;
XX
DT 08-SEP-1999 (first entry)
XX
DE Human SIGIRR protein.
XX
KW SIGIRR; anti-inflammatory; anti-autoimmune disease; chromosome 11;
KW interleukin-1 receptor; 11p15.5; arthrogryposis multiplex congenita;
KW breast cancer; insulin-dependent diabetes; sickle cell anaemia;
KW bladder cancer; detection; cell-signal transduction; gene therapy;
KW immune regulation; cell proliferation; cell death; cell migration;
KW cell interaction; inflammation; marker; mass spectrometry; human.
XX
OS Homo sapiens.
XX
PN WO9932626-A1.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98WO-US27368.
XX
PR 23-DEC-1997; 97US-0068770.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Sims JE;
XX
DR WPI; 1999-418925/35.
DR N-PSDB; AAX88091.

XX
PT New interleukin-1 receptor analog SIGIRR nucleic acid and proteins
XX
PS Claim 2; Page 69-71; 72pp; English.
XX
CC This invention describes a novel human SIGIRR DNA, its allelic variants
CC or species homologues which have anti-inflammatory and anti-autoimmune
CC disease activity. SIGIRR is an analog of interleukin-1 receptor. The
CC products of the invention are used (a) as probes or primers for
CC identifying nucleic acid that encodes proteins with SIGIRR activity; (b)
CC to identify human chromosome 11, to map genes on this chromosome and to
CC identify disease-related genes (particularly in the region 11p15.5 where
CC genes are present associated with e.g. arthrogryposis multiplex
CC congenita, breast cancer, insulin-dependent diabetes, sickle cell
CC anaemia, bladder cancer), including detection of defective genes; (c)
CC to study cell-signal transduction and the SIGIRR system, and (d) in
CC gene therapy. Sense and antisense oligonucleotides can be used to inhibit
CC expression of the SIGIRR gene. The proteins of the invention are used:
CC (i) to study cellular processes (immune regulation, proliferation, death,
CC migration, interaction with other cells and inflammation); (ii) to
CC identify and purify proteins that associate with SIGIRR ligands and
CC receptors, and to measure their biological activity; (iii) in screening
CC for, and rational design of, potential inhibitors of activity; (iv)
CC therapeutically against diseases mediated by SIGIRR polypeptide
CC counter-structures; (v) as molecular weight (m.w.) markers in
CC electrophoresis; (vi) for determining isoelectric points of unknown
CC proteins; (vii) as controls for determining the extent of protein
CC fragmentation (e.g. to aid characterization of protein structures by
CC mass spectrometry); (viii) for generation of antibodies (Ab); (ix) to
CC deliver diagnostic or therapeutic agents to cells that express SIGIRR
CC binding molecules.
XX
SQ Sequence 410 AA;

Query Match 100.0%; Score 626; DB 20; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.6e-57;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60
Db 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60

Qy 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
Db 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118

RESULT 2
AAB61139
ID AAB61139 standard; Protein; 410 AA.
XX
AC AAB61139;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human NOV9 protein.
XX
KW Human; NOVX; antiinflammatory; cytostatic; neuroprotective;
KW cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN WO200075321-A2.
XX
PD 14-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15303.
XX
PR 03-JUN-1999; 99US-0137322.
PR 16-MAR-2000; 2000US-0189810.
PR 22-MAR-2000; 2000US-0191158.
PR 30-MAR-2000; 2000US-0193086.
PR 31-MAY-2000; 2000US-0137322.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
XX
DR WPI; 2001-102403/11.
DR N-PSDB; AAF27857.
XX
PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
PT diagnostic marker, protein therapeutic, antibody or small molecule drug
PT target for treating immune, proliferative and metabolic diseases and
PT wound healing
XX
PS Claim 1; Page 36-38; 194pp; English.
XX
CC The present sequence is a new isolated polypeptide (NOVX). The NOVX
CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC treating or preventing NOVX-associated disorders. They are also useful
CC for determining the presence of or a predisposition to a disease
CC associated with altered levels of the NOVX polypeptide or nucleic acid.
CC These NOVX-associated disorders include hyperplasias, tumours,
CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC are especially useful in gene therapy. Specifically, NOVX is useful as
CC a diagnostic marker or prognostic marker, protein therapeutic and
CC antibody target or small molecule drug target to treat disorders in the
CC immune response pathway, thyroid and metabolic diseases, bone metabolic
CC disorders, diseases of the pancreas (e.g. diabetes or digestive
CC disorders), proliferative diseases, or tissue regeneration and
CC development (e.g. wound healing or treatment of burns).
XX
SQ Sequence 410 AA;

Query Match 100.0%; Score 626; DB 22; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.6e-57;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60
Db 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60

Qy 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
Db 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118

RESULT 3
AAU86151
ID AAU86151 standard; Protein; 504 AA.
XX
AC AAU86151;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human PRO342 polypeptide.
XX
KW Human; PRO; benign tumour; malignant tumour; lymphoid malignancy;
KW leukaemia; neuronal disorder; stromal disorder; blastocoelic disorder;
KW inflammatory disorder; immune disorder; angiogenic disorder;
KW cytostatic; neuroprotective.
XX
OS Homo sapiens.
XX
PN WO200153486-A1.
XX
PD 26-JUL-2001.
XX
PF 11-FEB-2000; 2000WO-US03565.
XX
PR 08-MAR-1999; 99WO-US05028.

PR 11-MAR-1999; 99US-123972P.
PR 11-MAY-1999; 99US-133459P.
PR 02-JUN-1999; 99WO-US12252.
PR 22-JUN-1999; 99US-140650P.
PR 22-JUN-1999; 99US-140653P.
PR 20-JUL-1999; 99US-144758P.
PR 26-JUL-1999; 99US-145698P.
PR 28-JUL-1999; 99US-146222P.
PR 17-AUG-1999; 99US-149395P.
PR 31-AUG-1999; 99US-151689P.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 05-JAN-2000; 2000WO-US00219.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;
PI Marsters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;
PI Watanabe CK, Wood WI;
XX
DR WPI; 2002-205567/26.
DR N-PSDB; ABK40277.
XX
PT Thirty five nucleic acids encoding PRO polypeptides, useful for
PT treating benign or malignant tumours, leukaemias and lymphoid
PT malignancies, inflammatory, angiogenic and immunologic disorders -
XX
PS Claim 61; Fig 48; 302pp; English.
XX
CC The present invention relates to the isolation of novel human PRO
CC polypeptides and the polynucleotide sequences encoding them. The
CC PRO polypeptides, agonists, antagonists or anti-PRO antibodies are
CC useful for treating benign or malignant tumours (e.g. renal, kidney,
CC bladder, breast, etc), leukaemias and lymphoid malignancies, other
CC disorders such as neuronal, glial, astrocytal, hypothalamic, glandular,
CC macrophagal, stromal and blastocoelic disorders, inflammatory, immune
CC and angiogenic disorders. The polynucleotide sequences are also
CC useful in gene therapy. AAU86128-AAU86162 represent the human PRO
CC polypeptides of the invention.
XX
SQ Sequence 504 AA;

Query Match 100.0%; Score 626; DB 23; Length 504;
Best Local Similarity 100.0%; Pred. No. 2e-57;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60

Qy 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118

RESULT 4
AAU17408
ID AAU17408 standard; Protein; 407 AA.
XX
AC AAU17408;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 973.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR N-PSDB; AAS27325.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID No 973; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 38.3%; Score 240; DB 22; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.5e-17;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 67 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 115

RESULT 5
AAM43652
ID AAM43652 standard; Protein; 407 AA.
XX
AC AAM43652;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 330.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
PN WO200155308-A2.
XX

PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488781/53.

DR N-PSDB; AAI63958.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Claim 11; SEQ ID NO 330; 664pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI63803-AAI64012) and
CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 407 AA;

Query Match 38.3%; Score 240; DB 22; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.5e-17;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 67 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 115

RESULT 6
AAM43583
ID AAM43583 standard; Protein; 410 AA.
XX
AC AAM43583;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 261.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
PN WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488781/53.
DR N-PSDB; AAI63889.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Claim 11; SEQ ID NO 261; 664pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI63803-AAI64012) and
CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 410 AA;

Query Match 38.3%; Score 240; DB 22; Length 410;
Best Local Similarity 100.0%; Pred. No. 7.6e-17;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 70 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118

RESULT 7
AAB58416
ID AAB58416 standard; Protein; 363 AA.
XX
AC AAB58416;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polypeptide sequence SEQ ID 754.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR N-PSDB; AAF18292.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 11; Page 1271-1272; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other

```
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   363 AA;

  Query Match             37.4%;  Score 234;  DB 21;  Length 363;
  Best Local Similarity   98.0%;  Pred. No. 2.8e-16;
  Matches   48;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      70 VKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
           |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      23 VKANLSEVLVYSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 71


RESULT 8
[redacted]30570
ID   AAB30570 standard; Protein; 871 AA.
XX
AC   AAB30570;
XX
DT   19-MAR-2001  (first entry)
XX
DE   A splice variant of a signal transduction polypeptide.
XX
KW   Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW   congestive heart failure; dilated congestive cardiomyopathy;
KW   hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW   mitral valve disease; aortic valve disease; tricuspid valve disease;
KW   myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW   atherosclerosis; cardiac tumour; microbial infection; splice variant.
XX
OS   Homo sapiens.
XX
PN   WO200063381-A1.
XX
PD   26-OCT-2000.
XX
PF   11-APR-2000; 2000WO-US09488.
XX
PR   16-APR-1999;   99US-0129553.
XX
PA   (SCIO-) SCIOS INC.
XX
[redacted]   Zeng W,  Stanton L,  Kong H;
[redacted]   WPI; 2001-007013/01.
XX
PT   Novel h19G5 polypeptides capable of regulating signal transduction and
PT   exhibiting kinase activity useful for identifying antibodies to treat
PT   cardiac diseases, and additional mediators of signal transduction -
XX
PS   Claim 1; Page 74-76; 81pp; English.
XX
CC   The present sequence represents a splice variant of human in signal
CC   transduction polypeptide. The polypeptide is designated H19G5. The
CC   protein is capable of regulating signal transduction and exhibits kinase
CC   activity. The H19G5 transcript is expressed in the heart. H19G5
CC   polypeptides and polynucleotides are useful for preventing or treating a
CC   cardiac disease, such as congestive heart failure, dilated congestive
CC   cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC   mitral valve disease, aortic valve disease or tricuspid valve disease,
CC   angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC   arterial or rhenovascular hypertension, arteriosclerosis, atherosclerosis
CC   and cardiac tumours in humans. The polypeptide is also useful for
CC   detecting the expression of a protein capable of regulating signal
CC   transduction or the expression of a protein capable of acting as a donor
CC   or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC   be used as probes for detecting discrete antigens expressed by tissue or
CC   cell samples, and therefore used in humans for localization and
CC   monitoring of microbial infection.
XX
SQ   Sequence   871 AA;

  Query Match             16.8%;  Score 105;  DB 22;  Length 871;
  Best Local Similarity   31.4%;  Pred. No. 0.034;
  Matches   33;  Conservative   11;  Mismatches   29;  Indels   32;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
           |||| ||    |: : :  || || | |          : |: |  | |||  ||
Db     363 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 406

Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
                 : | ||:|:      :| : |   |  | :|||: |
Db     407 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 443


RESULT 9
AAE16274
ID   AAE16274 standard; Protein; 871 AA.
XX
AC   AAE16274;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human kinase PKIN-20 protein.
XX
KW   Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;
KW   immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;
KW   Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;
KW   allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;
KW   autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;
KW   Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;
KW   rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;
KW   hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;
KW   cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;
KW   congestive heart failure; ischaemic heart disease; lung tumour; gout;
KW   fatty liver; Niemann-Pick's disease; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          575..827
FT                   /note= "Eukaryotic protein kinase domain"
FT   Domain          580..812
FT                   /label= Protein_kinase_domain
XX
PN   WO200196547-A2.
XX
PD   20-DEC-2001.
XX
PF   14-JUN-2001; 2001WO-US19444.
XX
PR   15-JUN-2000; 2000US-212073P.
PR   23-JUN-2000; 2000US-213467P.
PR   30-JUN-2000; 2000US-215651P.
PR   07-JUL-2000; 2000US-216605P.
PR   13-JUL-2000; 2000US-218372P.
PR   25-AUG-2000; 2000US-228056P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H,  Lal P,  Bandman O,  Borowsky ML,  Au-Young J,  Lu Y;
PI   Gandhi AR,  Tribouley CM,  Walia NK,  Yao MG,  Lu DAM,  Greenwald SR;
PI   Ramkumar J,  Griffin JA,  Kearney L,  Burford N,  Nguyen DB,  Tang YT;
PI   Baughn MR,  He A,  Thornton M,  Hafalia A,  Patterson C,  Gururajan R;
PI   Lo TP,  Khan F,  Recipon SA,  Azimzai Y,  Policky JL,  Ding L;
PI   Grether M,  Elliott VS,  Thangavelu K,  Batra S,  Ison CH;
XX
DR   WPI; 2002-090207/12.
DR   N-PSDB; AAD26467.
```

XX
PT New polypeptides, useful for diagnosing, treating or preventing
PT disorders of growth and development, cardiovascular and lipid, and
PT diseases such as cancer, comprise human kinase polypeptides -
XX
PS Claim 1; Page 164-165; 197pp; English.
XX
CC The invention relates to human kinase PKIN proteins and their
CC corresponding cDNAs. A composition containing PKIN agonist is useful for
CC treating a disease or condition associated with decreased expression of
CC PKIN and a composition comprising PKIN antagonist is useful for treating
CC a disease or condition associated with overexpression of PKIN. The
CC disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,
CC myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder
CC (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,
CC atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,
CC autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes
CC mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease,
CC osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome,
CC rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis,
CC bacterial, parasitic, fungal, viral, protozoal and helminthic infections)
CC growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,
CC Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio
CC vascular disease (arteriovenous fistula, hypertension, vasculitis,
CC aneurysms, congestive heart failure, angina pectoris, myocarditis,
CC ischaemic heart disease, chronic bronchitis, lung tumours); lipid
CC disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,
CC hypocholesterolaemia, obesity). PKIN DNA is useful for assessing
CC toxicity of a test compound and in gene therapy. The present sequence
CC is human PKIN-20 protein.
XX
SQ Sequence 871 AA;

Query Match 16.8%; Score 105; DB 23; Length 871;
Best Local Similarity 31.4%; Pred. No. 0.034;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

Qy 6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
Db 363 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 406

Qy 64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
Db 407 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 443

RESULT 10
AAB30567
ID AAB30567 standard; Protein; 1351 AA.
XX
AC AAB30567;
XX
DT 19-MAR-2001 (first entry)
XX
DE Amino acid sequence of a human signal transduction polypeptide.
XX
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW congestive heart failure; dilated congestive cardiomyopathy;
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection.
XX
OS Homo sapiens.
XX
PN WO200063381-A1.
XX
PD 26-OCT-2000.
XX
PF 11-APR-2000; 2000WO-US09488.
XX
PR 16-APR-1999; 99US-0129553.
XX
PA (SCIO-) SCIOS INC.
XX
PI Zeng W, Stanton L, Kong H;
XX
DR WPI; 2001-007013/01.
DR N-PSDB; AAC62285.
XX
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction -
XX
PS Claim 1; Page 55-57; 81pp; English.
XX
CC The present sequence represents a human protein with putative function
CC in signal transduction. The polypeptide is designated H19G5. The protein
CC is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC arterial or rhenovascular hypertension, arteriosclerosis, atherosclerosis
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection.
XX
SQ Sequence 1351 AA;

Query Match 16.8%; Score 105; DB 22; Length 1351;
Best Local Similarity 31.4%; Pred. No. 0.06;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

Qy 6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
Db 843 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 886

Qy 64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
Db 887 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 923

RESULT 11
AAB30568
ID AAB30568 standard; Protein; 1610 AA.
XX
AC AAB30568;
XX
DT 19-MAR-2001 (first entry)
XX
DE A full length human signal transduction polypeptide.
XX
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW congestive heart failure; dilated congestive cardiomyopathy;
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection.
XX
OS Homo sapiens.
XX
PN WO200063381-A1.
XX
PD 26-OCT-2000.
XX
PF 11-APR-2000; 2000WO-US09488.
XX
PR 16-APR-1999; 99US-0129553.
XX

PA (SCIO-) SCIOS INC.
XX
PI Zeng W, Stanton L, Kong H;
XX
DR WPI; 2001-007013/01.
DR N-PSDB; AAC62286.
XX
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction -
XX
PS Claim 1; Page 61-65; 81pp; English.
XX
CC The present sequence represents a human protein with putative function
CC in signal transduction. The polypeptide is designated H19G5. The protein
CC is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC [redacted] mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC arterial or rhenovascular hypertension, arteriosclerosis, atherosclerosis
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection.
XX
SQ Sequence 1610 AA;

Query Match 16.8%; Score 105; DB 22; Length 1610;
Best Local Similarity 31.4%; Pred. No. 0.075;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

```
Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
           |||| ||    |: :   || || | |          : |: |  | ||| ||
Db    1102 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 1145

Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
             : | ||:|:      :| : |   |  | :|||: |
Db    1146 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 1182
```

RESULT 12
AAB85504
ID AAB85504 standard; protein; 1618 AA.
XX
[redacted] AAB85504;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human protein kinase SGK145.
XX
KW Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human;
KW antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;
KW analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;
KW antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;
KW osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;
KW vasotropic; antidiabetic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155356-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02337.
XX
PR 25-JAN-2000; 2000US-0178078.
PR 31-JAN-2000; 2000US-0179364.
PR 17-FEB-2000; 2000US-0183173.
PR 17-MAR-2000; 2000US-0190162.
PR 29-MAR-2000; 2000US-0193404.
PR 13-NOV-2000; 2000US-0247013.
XX
PA (SUGE-) SUGEN INC.
XX
PI Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX
DR WPI; 2001-476202/51.
DR N-PSDB; AAH46904.
XX
PT Kinase polypeptides useful for treating cancers, Alzheimer's disease,
PT viral infections, diabetes, obesity, organ transplant rejection and
PT rheumatoid arthritis -
XX
PS Claim 7; Page 215; 218pp; English.
XX
CC The invention provides human protein kinases and protein kinase-like
CC enzymes and polynucleotides encoding the polypeptides. The kinase
CC polypeptides and their modulators are useful for treating a disease or
CC disorder such as cancer, immune-related diseases, cardiovascular disease,
CC brain or neuronal-associated disease and metabolic disorders, including
CC cancers of tissues, cancers of hematopoietic origin, diseases of the
CC central nervous system, diseases of the peripheral nervous system,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
CC lateral sclerosis, viral infections, infections caused by prions,
CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,
CC mood disorders, attention disorders, cognition disorders, hypotension,
CC hypertension, psychotic disorders, neurological disorders, dyskinesias,
CC metabolic disorders, and organ transplant rejection. They are also useful
CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,
CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic
CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders
CC such as diabetes, obesity, cardiovascular diseases such as reperfusion
CC injury, coronary thrombosis, clotting disorders and atherosclerosis,
CC ocular diseases such as glaucoma, retinopathy and macular degeneration,
CC psychiatric and neurological disorders such as anxiety, schizophrenia,
CC dementia, manic depression, etc. The polynucleotides are useful in gene
CC therapy techniques to treat the above mentioned disorders. Sequences
CC AAB85491-85522 represent the human protein kinases of the invention.
XX
SQ Sequence 1618 AA;

Query Match 16.8%; Score 105; DB 22; Length 1618;
Best Local Similarity 31.4%; Pred. No. 0.075;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

```
Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
           |||| ||    |: :   || || | |          : |: |  | ||| ||
Db    1110 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 1153

Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
             : | ||:|:      :| : |   |  | :|||: |
Db    1154 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 1190
```

RESULT 13
AAE24151
ID AAE24151 standard; Protein; 1665 AA.
XX
AC AAE24151;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human kinase (PKIN)-22 protein.
XX
KW Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;
KW acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
KW asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;
KW development; hepatitis; cardiovascular; hypertension; drug screening;
KW myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;
KW fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;

KW hypercholesterolaemia; obesity; gene therapy; cytostatic; anti-hIV;
KW neuroprotective; hepatotropic; hypotensive; cardiant; nephrotropic;
KW hyperlipidaemia; enzyme.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 68..128
FT /note= "Immunoglobulin domain"
FT Domain 165..418
FT /note= "Eukaryotic protein kinase domain"
FT Domain 167..401
FT /note= "Protein kinase domain"
FT Domain 1174..1235
FT /note= "Immunoglobulin domain"
FT Domain 1369..1621
FT /note= "Eukaryotic protein kinase domain"
FT Domain 1372..1606
FT /note= "Protein kinase domain"
XX
PN WO200233099-A2.
XX
PD 25-APR-2002.
XX
PF 20-OCT-2001; 2001WO-US47728.
XX
PR 20-OCT-2000; 2000US-242410P.
PR 27-OCT-2000; 2000US-244068P.
PR 03-NOV-2000; 2000US-245708P.
PR 09-NOV-2000; 2000US-247672P.
PR 16-NOV-2000; 2000US-249565P.
PR 22-NOV-2000; 2000US-252730P.
PR 01-DEC-2000; 2000US-250807P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gururajan R, Baughn MR, Walia NK, Elliott VS, Xu Y, Arvizu C;
PI Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;
PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM;
PI Lal PG, Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;
PI Thangavelu K, Khan FA, Ison CH;
XX
DR WPI; 2002-454603/48.
DR N-PSDB; AAD38865.
XX
PT New human kinase polypeptide, for diagnosing, preventing and treating
PT cancer, immune system disorders, growth and development disorders,
PT cardiovascular disorders and lipid disorders -
XX
PS Claim 1; Page 182-186; 210pp; English.
XX
CC The invention relates human kinases (PKIN) and their corresponding
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
CC treating and preventing cancer, an immune system disorder (e.g.,
CC acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,
CC asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders
CC affecting growth and development (e.g., arteriosclerosis, cirrhosis,
CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial
CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty
CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,
CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
CC condition or a disease associated with the expression of PKIN in a
CC biological sample. A composition comprising PKIN or an agonist or
CC antagonist of PKIN is useful for treating a disease or condition
CC associated with decreased or increased expression of functional PKIN.
CC PKIN is useful in a number of drug screening techniques and to analyse
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
CC knockin humanised animals or transgenic animals to model human diseases,
CC and in somatic or germline gene therapy. The present sequence is human
CC PKIN protein.
XX
SQ Sequence 1665 AA;

Query Match 16.8%; Score 105; DB 23; Length 1665;
Best Local Similarity 31.4%; Pred. No. 0.078;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

Qy 6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
Db 1157 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 1200

Qy 64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
Db 1201 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 1237

RESULT 14
AAO15372
ID AAO15372 standard; Protein; 1665 AA.
XX
AC AAO15372;
XX
DT 19-SEP-2002 (first entry)
XX
DE Human myosin light chain kinase subfamily-related kinase protein.
XX
KW Human; gene therapy; chromosome 1; kinase protein;
KW myosin light chain kinase subfamily; kinase protein-mediated disease;
KW transgenic animal.
XX
OS Homo sapiens.
XX
PN WO200240683-A2.
XX
PD 23-MAY-2002.
XX
PF 22-OCT-2001; 2001WO-US32616.
XX
PR 14-NOV-2000; 2000US-0711134.
PR 17-MAY-2001; 2001US-0858664.
XX
PA (PEKE) PE CORP NY.
XX
PI Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-500223/53.
DR N-PSDB; AAL43908, AAL43909.
XX
PT New kinase proteins related to myosin light chain kinase subfamily and
PT encoding polynucleotide, useful for diagnosing, treating disease or
PT condition mediated by the kinase protein and for identifying modulators
PT -
XX
PS Claim 1; Fig 2; 96pp; English.
XX
CC The invention comprises the amino acid and coding sequences (located on
CC chromosome 1) of a human kinase protein that is related to the myosin
CC light chain kinase subfamily. The human kinase DNA and protein sequences
CC of the invention are useful for identifying agents that modulate the
CC activity of the human kinase protein. Kinase-modulating agents are useful
CC for treating a disease or condition mediated by a human kinase protein.
CC The human kinase DNA sequences can be used to produce transgenic animals
CC which are useful for studying the function of kinase proteins and
CC identifying/evaluating modulators of kinase protein activity. The present
CC amino acid sequence represents the human kinase protein of the invention.
XX
SQ Sequence 1665 AA;

Query Match 16.8%; Score 105; DB 23; Length 1665;
Best Local Similarity 31.4%; Pred. No. 0.078;
Matches 33; Conservative 11; Mismatches 29; Indels 32; Gaps 5;

Qy 6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
Db 1157 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 1200

```
Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
                 : | ||:|:      :| : |   |  | :|||: |
Db    1201 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 1237

RESULT 15
AAB30569
ID   AAB30569 standard; Protein; 2596 AA.
XX
AC   AAB30569;
XX
DT   19-MAR-2001  (first entry)
XX
DE   A splice variant of a signal transduction polypeptide.
XX
KW   Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW   congestive heart failure; dilated congestive cardiomyopathy;
KW   hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW   mitral valve disease; aortic valve disease; tricuspid valve disease;
KW   myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW   atherosclerosis; cardiac tumour; microbial infection; splice variant.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          325..504
FT                   /note= "guanine nucleotide exchange factor domain"
FT   Domain          1094..1351
FT                   /note= "kinase domain"
FT   Domain          2301..2553
FT                   /note= "kinase domain"
XX
PN   WO200063381-A1.
XX
PD   26-OCT-2000.
XX
PF   11-APR-2000; 2000WO-US09488.
XX
PR   16-APR-1999;   99US-0129553.
XX
PA   (SCIO-) SCIOS INC.
XX
PI   Zeng W,  Stanton L,  Kong H;
XX
DR   WPI; 2001-007013/01.
DR   N-PSDB; AAC62287.
XX
PT   Novel h19G5 polypeptides capable of regulating signal transduction and
PT   exhibiting kinase activity useful for identifying antibodies to treat
PT   cardiac diseases, and additional mediators of signal transduction -
XX
PS   Claim 1; Page 68-74; 81pp; English.
XX
CC   The present sequence represents a splice variant of human in signal
CC   transduction polypeptide. The polypeptide is designated H19G5. The
CC   protein is capable of regulating signal transduction and exhibits kinase
CC   activity. The H19G5 transcript is expressed in the heart. H19G5
CC   polypeptides and polynucleotides are useful for preventing or treating a
CC   cardiac disease, such as congestive heart failure, dilated congestive
CC   cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC   mitral valve disease, aortic valve disease or tricuspid valve disease,
CC   angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC   arterial or rhenovascular hypertension, arteriosclerosis, atherosclerosis
CC   and cardiac tumours in humans. The polypeptide is also useful for
CC   detecting the expression of a protein capable of regulating signal
CC   transduction or the expression of a protein capable of acting as a donor
CC   or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC   be used as probes for detecting discrete antigens expressed by tissue or
CC   cell samples, and therefore used in humans for localization and
CC   monitoring of microbial infection.
XX
SQ   Sequence   2596 AA;
```

```
  Query Match             16.8%;  Score 105;  DB 22;  Length 2596;
  Best Local Similarity   31.4%;  Pred. No. 0.14;
  Matches   33;  Conservative   11;  Mismatches   29;  Indels   32;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
           |||| ||    |: :   || || | |          : |: |  | ||| ||
Db    2088 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 2131

Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
                 : | ||:|:      :| : |   |  | :|||: |
Db    2132 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 2168


Search completed: November 14, 2002, 17:30:01
Job time : 38 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 14, 2002, 17:29:38 ; Search time 17 Seconds
(without alignments)
667.286 Million cell updates/sec

Title: US-09-598-443-2_COPY_1_118
Perfect score: 626
Sequence: 1 MPGVCDRAPDFLSPSEDQVL.....SIQNISFSSFTLQRAGPTSH 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	93.5	14.9	7962	2	I38346	elastic titin - hu
2	91.5	14.6	1323	2	PN0568	connectin 3B - chi
3	90	14.4	6831	2	A88852	protein unc-22 [im
4	90	14.4	6839	2	S57242	twitchin [similari
5	90	14.4	7160	2	T27935	hypothetical prote
6	87.5	14.0	5175	2	T20992	hypothetical prote
7	87.5	14.0	5198	2	T43290	hemicentin precurs
8	83.5	13.3	244	2	AE3507	amidotransferase h
9	83.5	13.3	1694	2	S50065	sialoadhesin - mou
10	82	13.1	588	2	I37202	B-CAM protein - hu
11	82	13.1	628	2	I38000	Lutheran blood gro
12	81.5	13.0	650	1	JC1450	fibroblast growth
13	81	12.9	238	2	A49633	Ig lambda-like cha
14	80.5	12.9	799	2	S18209	fibroblast growth
15	80.5	12.9	6805	2	S20901	titin - rabbit (fr
16	80	12.8	1896	2	T08851	Down syndrome cell
17	80	12.8	6642	2	T29757	protein UNC-89 - C
18	80	12.8	26926	1	I38344	titin, cardiac mus
19	78.5	12.5	483	2	T17346	hypothetical prote
20	78	12.5	2783	2	T34416	hypothetical prote
21	77.5	12.4	361	2	PN0020	fibroblast growth
22	77.5	12.4	822	1	TVMSFG	fibroblast growth
23	77.5	12.4	832	2	JH0393	fibroblast growth
24	77.5	12.4	1091	2	A58532	glial cell membran
25	77	12.3	733	2	I49293	fibroblast growth
26	77	12.3	822	2	I49289	fibroblast growth
27	77	12.3	889	2	E87304	TonB-dependent rec
28	77	12.3	1240	2	T03097	CDO protein - huma
29	76.5	12.2	1142	2	S36845	myosin-binding pro
30	76.5	12.2	1241	2	T37190	nephrin - human
31	76	12.1	602	2	A35564	prostaglandin-endo
32	75.5	12.1	662	2	C40862	heparin-binding gr
33	75.5	12.1	729	2	A56795	fibroblast growth
34	75.5	12.1	822	1	TVHUFG	fibroblast growth
35	75	12.0	276	2	S75249	esterase sll0992 -
36	75	12.0	480	2	A56182	fibroblast growth
37	75	12.0	824	2	S24108	protein-tyrosine k
38	74.5	11.9	802	1	TVHUF4	fibroblast growth
39	74.5	11.9	819	1	TVCHFG	fibroblast growth
40	74	11.8	162	2	I51668	tumor suppressor -
41	73.5	11.7	602	2	S39782	cyclooxygenase 1 -
42	73.5	11.7	1021	2	T42634	connectin/titin -
43	73.5	11.7	1666	2	A48594	skelemin - mouse
44	73	11.7	750	2	S41051	fibroblast growth
45	73	11.7	1272	2	S26180	neurofascin - chic

ALIGNMENTS

RESULT 1
I38346
elastic titin - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38346
R;Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:96026330; PMID:7569978
A;Accession: I38346
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-7962 <RES>
A;Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q31

Query Match 14.9%; Score 93.5; DB 2; Length 7962;
Best Local Similarity 32.8%; Pred. No. 4.3;
Matches 38; Conservative 7; Mismatches 32; Indels 39; Gaps 7;

```
Qy      10 DFLSPSEDQVLR----------------PALGSSVALNCTAWVVSGPHCSLP-SVQWL 50
           || | | |  ||                ||||: : |    |||   ||| | ||
Db    1883 DFGSSSCDAYLRVLDQNIPPSFTKKLTKMDKVLGSSIHMECK---VSG---SLPISAQWF 1936

Qy      51 KDGLPLGIGGHYSL--HEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNIS 104
           |||  :     | |  || |          ||  | ||    |    :|| : |::
Db    1937 KDGKEISTSAKYRLVCHERS---------VS--LEVNNLELEDTANYTCKVSNVA 1980
```

RESULT 2
PN0568
connectin 3B - chicken (fragment)
N;Alternate names: Cn3B protein
C;Species: Gallus gallus (chicken)
C;Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C;Accession: PN0568
R;Maruyama, K.; Endo, T.; Kume, H.; Kawamura, Y.; Kanzawa, N.; Nakauchi, Y.; Kimura,
Biochem. Biophys. Res. Commun. 194, 1288-1291, 1993
A;Title: A novel domain sequence of connectin localized at the I band of skeletal mus
A;Reference number: PN0568; MUID:93356802; PMID:8352787
A;Accession: PN0568
A;Molecule type: mRNA
A;Residues: 1-1323 <MAR>
A;Cross-references: DDBJ:D16541; NID:g391629; PID:d1004495; PID:g391630
A;Experimental source: skeletal muscle
C;Comment: This protein string-like single molecule spans from the Z line to the M li

Query Match 14.6%; Score 91.5; DB 2; Length 1323;

```
  Best Local Similarity   24.5%;  Pred. No. 0.83;
  Matches   27;  Conservative   21;  Mismatches   31;  Indels   31;  Gaps    5;

Qy      19 VLRPAL----------GSSVALNCTAWVVSGPHCSLPSV--QWLKDGLPLGIGGHYSLH 65
           |: ||:          |::  |:|          |:|:   || ::| |: ||| ||
Db     959 VIEPAVERHLHDTTFKEGNTCTLSC--------QFSIPNAKSQWYRNGRPIKIGGRYS-- 1008

Qy      66 EYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGP 115
             : |   : :::: |       ||  | :|| : |:  ::     | |
Db    1009 --TQVSDKVHKLIIKDV------RTEDQGQYTCKLDNLETTADLTIEAEP 1050


RESULT 3
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: A88852
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88852
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-6831 <STO>
A;Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00022
C;Genetics:
A;Gene: unc-22
A;Map position: 4
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

  Query Match             14.4%;  Score 90;  DB 2;  Length 6831;
  Best Local Similarity   24.0%;  Pred. No. 8;
  Matches   23;  Conservative   21;  Mismatches   40;  Indels   12;  Gaps    3;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           | | |:   : :::    |: : : |          | |  :|:|||:|| :|| |
Db     101 RGPSFV--GKPRIIPKDGGALIVMECKV-----KSASTPVAKWMKDGVPLSMGGLYHA-- 151

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
              : ::| :      | :   |:   | : |:|:|
Db     152 ---IFSDLGDQTYLCQLEIRGPSSSDAGQYRCNIRN 184


RESULT 4
S57242
[illegible]chin [similarity] - Caenorhabditis elegans
[illegible]lternate names: myosin-regulating protein
N;Contains: protein kinase (EC 2.7.1.-)
C;Species: Caenorhabditis elegans
C;Date: 28-Oct-1995 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C;Accession: S57242; S07571; S06797; S57218; T27934; T28030
R;Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
submitted to the EMBL Data Library, February 1993
A;Description: Additional sequence complexity within twitching of Caenorhabditis elegans
A;Reference number: S57242
A;Accession: S57242
A;Molecule type: DNA
A;Residues: 1-6839 <BEN1>
A;Cross-references: EMBL:L10351
A;Experimental source: var. Bristol
R;Benian, G.
submitted to the EMBL Data Library, November 1989
A;Reference number: S07571
A;Accession: S07571
A;Molecule type: DNA
A;Residues: 792-6839 <BEN2>
A;Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898
A;Experimental source: var. Bristol
R;Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989
A;Title: Sequence of an unusually large protein implicated in regulation of myosin ac
A;Reference number: S06797; MUID:90044042; PMID:2812002
A;Accession: S06797
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 806-1175;1178-1998,'Y',2000-3040,'I',3042-3335,'I',3337-5693;5696-6359,'I
A;Cross-references: EMBL:X15423
A;Experimental source: var. Bristol
R;Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
Genetics 134, 1097-1104, 1993
A;Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded p
A;Reference number: S57218; MUID:93387664; PMID:8397135
A;Accession: S57218
A;Molecule type: DNA
A;Residues: 2-99;108-194,'Q',196-206;374-468;658-753 <BEN4>
A;Experimental source: var. Bristol
R;White, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20442
A;Accession: T27934
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'MGIPGKKCKQ',19-6839 <WIL>
A;Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a
A;Experimental source: clone ZK617
R;Harris, B.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20458
A;Accession: T28030
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'MGIPGKKCKQ',19-6839 <WI2>
A;Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a
A;Experimental source: clone ZK829
C;Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C;Genetics:
A;Gene: unc-22; CESP:ZK617.1a
A;Map position: 4
A;Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3;
152/3; 6691/3; 6776/1; 6808/3
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homolog
C;Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine
F;806-898,899-990,991-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-245
96-5790,6263-6356,6386-6478,6541-6635,6649-6742,6745-6838/Region: motif 2
F;1274-1372,1373-1473,1568-1670,1671-1769,1865-1964,1965-2065,2159-2258,2259-2357,245
23,4215-4313,4314-4415,4416-4516,4612-4710,4711-4811,4908-5009,5010-5109,5110-5210,53
F;5940-6197/Domain: protein kinase homology <KIN>
F;5948-5956/Region: protein kinase ATP-binding motif
F;5971/Active site: Lys #status predicted

  Query Match             14.4%;  Score 90;  DB 2;  Length 6839;
  Best Local Similarity   24.0%;  Pred. No. 8;
  Matches   23;  Conservative   21;  Mismatches   40;  Indels   12;  Gaps    3;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           | | |:   : :::    |: : : |          | |  :|:|||:|| :|| |
Db     109 RGPSFV--GKPRIIPKDGGALIVMECKV-----KSASTPVAKWMKDGVPLSMGGLYHA-- 159

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
              : ::| :      | :   |:   | : |:|:|
Db     160 ---IFSDLGDQTYLCQLEIRGPSSSDAGQYRCNIRN 192


RESULT 5
T27935
hypothetical protein ZK617.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C;Accession: T27935; T28031
R;White, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20442
```

```
A;Accession: T27935
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7160 <WIL>
A;Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b
A;Experimental source: clone ZK617
R;Harris, B.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20458
A;Accession: T28031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7160 <WI2>
A;Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b
A;Experimental source: clone ZK829
C;Genetics:
A;Gene: CESP:ZK617.1b
A;Map position: 4
A;Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 59
7/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

  Query Match             14.4%;  Score 90;  DB 2;  Length 7160;
  Best Local Similarity   24.0%;  Pred. No. 8.4;
  Matches   23;  Conservative   21;  Mismatches   40;  Indels   12;  Gaps    3;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           | | |:   : :::    |: : : |         | |  :|:|||:|| :|| |
Db     101 RGPSFV--GKPRIIPKDGGALIVMECKV-----KSASTPVAKWMKDGVPLSMGGLYHA-- 151

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
             : ::| :      | :   |:   | : |:|:|
Db     152 ---IFSDLGDQTYLCQLEIRGPSSSDAGQYRCNIRN 184


RESULT 6
T20992
hypothetical protein F15G9.4a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20992; T24733
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20992
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24733
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5175 <WI2>
A;Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: CESP:F15G9.4a
A;Map position: X
A;Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/

  Query Match             14.0%;  Score 87.5;  DB 2;  Length 5175;
  Best Local Similarity   28.1%;  Pred. No. 10;
  Matches   32;  Conservative   15;  Mismatches   38;  Indels   29;  Gaps    6;

Qy       1 MPG-------VCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDG 53
           |||       :   || |:   ::     |:|: | | |:|     ||   |:: | |||
Db    4555 MPGPRSSRTVLLHAAPQFIVKPKNTT--AAIGAIVELRCSA--AGPPH---PTITWAKDG 4607

Qy      54 LPLGIGGHYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSS 107
                          :: :  |:  |  | | : ||   | :||  ||   ||
Db    4608 --------------KLIEDSKFEIAYSH-LKVTLNSTSDSGEYTCMAQNSVGSS 4646


RESULT 7
T43290
hemicentin precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C;Accession: T43290; T20993; T24734
R;Vogel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A;Description: Hemicentin is required for hemidesmosome mediated cell adhesion and ge
A;Reference number: Z22396
A;Accession: T43290
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5198 <VOG>
A;Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R;Sulston, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19355
A;Accession: T20993
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WIL>
A;Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone F15G9
R;Kershaw, J.
submitted to the EMBL Data Library, December 1994
A;Reference number: Z19929
A;Accession: T24734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5198 <WI2>
A;Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b
A;Experimental source: clone T09B9
C;Genetics:
A;Gene: him-4; F15G9.4b
A;Map position: X
A;Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

  Query Match             14.0%;  Score 87.5;  DB 2;  Length 5198;
  Best Local Similarity   28.1%;  Pred. No. 10;
  Matches   32;  Conservative   15;  Mismatches   38;  Indels   29;  Gaps    6;

Qy       1 MPG-------VCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDG 53
           |||       :   || |:   ::     |:|: | | |:|     ||   |:: | |||
Db    4555 MPGPRSSRTVLLHAAPQFIVKPKNTT--AAIGAIVELRCSA--AGPPH---PTITWAKDG 4607

Qy      54 LPLGIGGHYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSS 107
                          :: :  |:  |  | | : ||   | :||  ||   ||
Db    4608 --------------KLIEDSKFEIAYSH-LKVTLNSTSDSGEYTCMAQNSVGSS 4646


RESULT 8
AE3507
amidotransferase hisH (EC 2.4.2.-) [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AE3507
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A;Reference number: AD3252; PMID:11756688
A;Accession: AE3507
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 1-244 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL53224.1; PID:g17984101; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI2043
A;Map position: I
C;Superfamily: amidotransferase hisH; trpG homology
C;Keywords: glycosyltransferase; pentosyltransferase

  Query Match             13.3%;  Score 83.5;  DB 2;  Length 244;
  Best Local Similarity   24.6%;  Pred. No. 0.71;
  Matches   33;  Conservative   13;  Mismatches   43;  Indels   45;  Gaps    6;

Qy       1 MPGV---------CDRAPDFLSPSEDQVL---RPALGSSVAL------------NCTAW 35
          :|||          |  |  :    | ||   || ||  | :             |   |
Db      78 LPGVGAYADCRRGLDAVPGMVEALNDTVLKKARPFLGICVGMQLMSERGLEKTVTNGLGW 137

Qy      36 VV---------SGPHCSLPSVQW----------LKDGLPLGIGG--HYSLHEYSWVKANLS 75
          :          |       :|  : |              : ||:| |  |   | :| |      | |
Db     138 IAGDVREMVPSDASLKIPQIGWNRIHVKHSHPIFDGIPTGDDGLHAYFVHSYMLDAKNAS 197

[redacted]  76 EVLVSSVLGVNVTS 89
          :||  :  | :|||:
Db     198 DVLAVTDYGGDVTA 211


RESULT 9
S50065
sialoadhesin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
C;Accession: S50065
R;Crocker, P.R.; Mucklow, S.; Bouckson, V.; McWilliam, A.; Willis, A.C.; Gordon, S.; Mil
EMBO J. 13, 4490-4503, 1994
A;Title: Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells
A;Reference number: S50065; MUID:95009950; PMID:7925291
A;Accession: S50065
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1694 <CRO>
A;Cross-references: EMBL:Z36293; NID:g557253; PIDN:CAA85290.1; PID:g557254

  Query Match             13.3%;  Score 83.5;  DB 2;  Length 1694;
  Best Local Similarity   21.6%;  Pred. No. 6.9;
  Matches   37;  Conservative   17;  Mismatches   42;  Indels   75;  Gaps    6;

Qy      12 LSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGH---------- 61
          || |   :|    |  | | |    |:  : :: :||| :||: ||: ||
Db     246 LSSSGRNILP---GDPVTLTCR---VNSSYPAVSAVQWARDGVNLGVTGHVLRLFSAAWN 299

[redacted]  62 --------------------YSLH--------------------------------EY 67
                             |||                                |
Db     300 DSGAYTCQATNDMGSLVSSPLSLHVFMAEVKMNPAGPVLENETVTLLCSTPKEAPQELRY 359

Qy      68 SWVKAN-LSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTS 117
          || | : | |   :| | :   :    | : | :||    |    :|: | |
Db     360 SWYKNHILLEDAHASTLHLPAVTRADTGFYFCEVQNAQGS----ERSSPLS 406


RESULT 10
I37202
B-CAM protein - human
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C;Accession: I37202; S47272
R;Campbell, I.G.; Foulkes, W.D.; Senger, G.; Trowsdale, J.; Garin-Chesa, P.; Rettig, W.J
Cancer Res. 54, 5761-5765, 1994
A;Title: Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers:
A;Reference number: I37202; MUID:95042297; PMID:7954395
A;Accession: I37202
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-588 <RES>
A;Cross-references: EMBL:X80026; NID:g535178; PIDN:CAA56327.1; PID:g535179
C;Genetics:
A;Gene: B-CAM

  Query Match             13.1%;  Score 82;  DB 2;  Length 588;
  Best Local Similarity   36.2%;  Pred. No. 2.8;
  Matches   29;  Conservative   12;  Mismatches   33;  Indels    6;  Gaps    3;

Qy      10 DFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSW 69
           | |  || :||   | ||  :|||:   | |      |:::| ||  ||| |   ||   ::
Db     362 DPLELSEGKVLSLPLNSSAVVNCS---VHG--LPTPALRWTKDSTPLGDGPMLSLSSITF 416

Qy      70 VKANLSEVLVSSVLGVNVTS 89
             :| : |  :|:  | | |
Db     417 -DSNGTYVCEASLPTVPVLS 435


RESULT 11
I38000
Lutheran blood group glycoprotein precursor - human
C;Species: Homo sapiens (man)
C;Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 05-Nov-1999
C;Accession: I38000; S51663
R;Parsons, S.F.; Mallinson, G.; Holmes, C.H.; Houlihan, J.M.; Simpson, K.L.; Mawby, W
Proc. Natl. Acad. Sci. U.S.A. 92, 5496-5500, 1995
A;Title: The Lutheran blood group glycoprotein, another member of the immunoglobulin
A;Reference number: I38000; MUID:95296337; PMID:7777537
A;Accession: I38000
A;Molecule type: mRNA
A;Residues: 1-628 <RES>
A;Cross-references: EMBL:X83425; NID:g603559; PIDN:CAA58449.1; PID:g603560
A;Note: parts of this sequence, including the amino end of the mature form, were conf
C;Genetics:
A;Gene: GDB:LU
A;Cross-references: GDB:120155; OMIM:111200
A;Map position: 19q12-19q13
C;Keywords: glycoprotein
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-628/Product: Lutheran blood group glycoprotein #status experimental <MAT>

  Query Match             13.1%;  Score 82;  DB 2;  Length 628;
  Best Local Similarity   36.2%;  Pred. No. 3;
  Matches   29;  Conservative   12;  Mismatches   33;  Indels    6;  Gaps    3;

Qy      10 DFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSW 69
           | |  || :||   | ||  :|||:   | |      |:::| ||  ||| |   ||   ::
Db     362 DPLELSEGKVLSLPLNSSAVVNCS---VHG--LPTPALRWTKDSTPLGDGPMLSLSSITF 416

Qy      70 VKANLSEVLVSSVLGVNVTS 89
             :| : |  :|:  | | |
Db     417 -DSNGTYVCEASLPTVPVLS 435


RESULT 12
JC1450
fibroblast growth factor receptor 4 - rat
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-2000
C;Accession: JC1450; PT0191
R;Horlick, R.A.; Stack, S.L.; Cooke, G.M.
Gene 120, 291-295, 1992
A;Title: Cloning, expression and tissue distribution of the gene encoding rat fibrobl
A;Reference number: JC1450; MUID:93013049; PMID:1398143
A;Accession: JC1450
A;Molecule type: mRNA
A;Residues: 1-650 <HOR>
A;Cross-references: GB:M91599; NID:g204137; PIDN:AAA41157.1; PID:g204138
R;Lai, C.; Lemke, G.
Neuron 6, 691-704, 1991
A;Title: An extended family of protein-tyrosine kinase genes differentially expressed
A;Reference number: PT0183; MUID:91222560; PMID:2025425
A;Accession: PT0191
```

```
A;Molecule type: mRNA
A;Residues: 465-518 <LAI>
A;Experimental source: sciatic nerve
C;Genetics:
A;Gene: FGFR4; tyro-9
C;Function:
A;Description: receptor mediating effects of fibroblast growth factor
A;Note: expressed in normal lung; expressed in some carcinomas
C;Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prote
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; growth factor receptor;
F;11-72/Domain: immunoglobulin homology <IM1>
F;110-181/Domain: immunoglobulin homology <IM2>
F;218-238/Domain: transmembrane #status predicted <TMM>
F;239-650/Domain: intracellular #status predicted <INT>
F;313-598/Domain: protein kinase homology <KIN>
F;321-329/Region: protein kinase ATP-binding motif
F;104,136,157,168/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;351,368,460/Active site: Lys, Glu, Asp #status predicted
F;[illegible]65,478/Binding site: magnesium (Asn, Asp) #status predicted
F;[illegible]91/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte

  Query Match             13.0%;  Score 81.5;  DB 1;  Length 650;
  Best Local Similarity   31.9%;  Pred. No. 3.6;
  Matches   29;  Conservative   11;  Mismatches   36;  Indels   15;  Gaps    4;

Qy      24 LGSSVALNCTAWVVSGPHCSLPSVQWLK----DGLPLGIGGHYSLHEYSWVKANLSEVLV 79
           :||:| | |  :  : ||     :|||     :|  ||  |   :           | |||
Db     109 VGSNVELLCKVYSDAQPH-----IQWLKHIVINGSSLGADGFPYVQVLKTTDINSSEV-- 161

Qy      80 SSVLGVNVTSTEVYGAFTCSIQN---ISFSS 107
             || :   | |  | :||   |   :|: |
Db     162 -EVLYLRNVSAEDAGEYTCLAGNSIGLSYQS 191


RESULT 13
A49633
Ig lambda-like chain, V-C region - nurse shark
C;Species: Ginglymostoma cirratum (nurse shark)
C;Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
C;Accession: A49633
R;Greenberg, A.S.; Steiner, L.; Kasahara, M.; Flajnik, M.F.
Proc. Natl. Acad. Sci. U.S.A. 90, 10603-10607, 1993
A;Title: Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein r
A;Reference number: A49633; MUID:94068449; PMID:8248152
A;Accession: A49633
A;Status: preliminary
A;[illegible]olecule type: mRNA
A;[illegible]esidues: 1-238 <GRE>
A;[illegible]xperimental source: spleen
A;Note: sequence extracted from NCBI backbone (NCBIN:141012, NCBIP:141013)
C;Superfamily: immunoglobulin V region; immunoglobulin homology
F;36-110/Domain: immunoglobulin homology <IMM>

  Query Match             12.9%;  Score 81;  DB 2;  Length 238;
  Best Local Similarity   29.3%;  Pred. No. 1.2;
  Matches   29;  Conservative   18;  Mismatches   36;  Indels   16;  Gaps    3;

Qy      12 LSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLP---SVQWLKDGLPLGIGGHYSLHEYS 68
           | || |||             || :|   :   |    ||| |||  :  |   |    :
Db     141 LPPSPDQVQTKG---------TATLVCLANHFYPDELQVQWKKDGAVISDGVQTS----N 187

Qy      69 WVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSS 107
           :::|: |   |||:| :: :  |    |:|:: ::: ||
Db     188 YLRASDSTYSVSSLLTLSGSDWESNARFSCALTHVTLSS 226


RESULT 14
S18209
fibroblast growth factor receptor 4 precursor (clone 61) - mouse
N;Alternate names: tyrosine kinase Mpk-11
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Mus musculus (house mouse)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 16-Jul-1999
C;Accession: S18209; S26751; S30497
R;Stark, K.L.; McMahon, J.A.; McMahon, A.P.
Development 113, 641-651, 1991
A;Title: FGFR-4, a new member of the fibroblast growth factor receptor family, expres
A;Reference number: S18209; MUID:92146274; PMID:1723680
A;Accession: S18209
A;Molecule type: mRNA
A;Residues: 1-799 <STA1>
A;Cross-references: EMBL:X59927
R;Stark, K.L.
submitted to the EMBL Data Library, May 1991
A;Reference number: S26751
A;Accession: S26751
A;Molecule type: mRNA
A;Residues: 1-485,'QVVRAEAFG',486-799 <STA2>
A;Cross-references: EMBL:X59927; NID:g50968; PIDN:CAA42551.1; PID:g50969
R;Gilardi-Hebenstreit, P.; Nieto, M.A.; Frain, M.; Mattei, M.G.; Chestier, A.; Wilkin
Oncogene 7, 2499-2506, 1992
A;Title: An Eph-related receptor protein tyrosine kinase gene segmentally expressed i
A;Reference number: S30496; MUID:93096484; PMID:1281307
A;Accession: S30497
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 611-667 <GIL>
A;Cross-references: EMBL:X57236; NID:g53187; PIDN:CAA40512.1; PID:g53188
C;Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; pr
C;Keywords: ATP; duplication; glycoprotein; growth factor receptor; phosphotransferas
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-799/Product: fibroblast growth factor receptor 4 #status predicted <MAT>
F;19-366/Domain: extracellular #status predicted <EXT>
F;162-223/Domain: immunoglobulin homology <IMM>
F;367-387/Domain: transmembrane #status predicted <TMM>
F;388-799/Domain: intracellular #status predicted <INT>
F;462-747/Domain: protein kinase homology <KIN>
F;470-478/Region: protein kinase ATP-binding motif
F;54-98,169-221,268-330/Disulfide bonds: #status predicted
F;500,517,609/Active site: Lys, Glu, Asp #status predicted

  Query Match             12.9%;  Score 80.5;  DB 2;  Length 799;
  Best Local Similarity   30.8%;  Pred. No. 5.7;
  Matches   28;  Conservative   11;  Mismatches   37;  Indels   15;  Gaps    4;

Qy      24 LGSSVALNCTAWVVSGPHCSLPSVQWLK----DGLPLGIGGHYSLHEYSWVKANLSEVLV 79
           :|| | | |  :  : ||     :||||    :|   |  |   :         |:|||
Db     260 VGSDVELLCKVYSDAQPH-----IQWLKHVVINGSSFGADGFPYVQVLKTTDINISEV-- 312

Qy      80 SSVLGVNVTSTEVYGAFTCSIQN---ISFSS 107
             || :   | |  | :||   |   :|: |
Db     313 -QVLYLRNVSAEDAGEYTCLAGNSIGLSYQS 342


RESULT 15
S20901
titin - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C;Accession: S20901; I46520
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380; PMID:1582406
A;Accession: S20901
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-6805 <LAB>
A;Cross-references: EMBL:X64696
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
R;Labeit, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U
Nature 345, 273-276, 1990
A;Title: A regular pattern of two types of 100-residue motif in the sequence of titin
A;Reference number: I46520; MUID:90238553; PMID:2129545
A;Accession: I46520
A;Status: translated from GB/EMBL/DDBJ
```

```
A;Molecule type: mRNA
A;Residues: 4235-5250 <LA2>
A;Cross-references: EMBL:X17329; NID:g1756; PIDN:CAA35207.1; PID:g930251
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prc
C;Keywords: muscle

  Query Match             12.9%;  Score 80.5;  DB 2;  Length 6805;
  Best Local Similarity   24.4%;  Pred. No. 70;
  Matches   21;  Conservative   16;  Mismatches   30;  Indels   19;  Gaps    3;

Qy      36 VVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGA 95
           |:  |    | : |:||| ||                 : |: |   |::| :  :| : :|
Db    6197 VIGRPR---PEIFWVKDGEPL----------RQTTRVNVEETATSTILHIKESSKDDFGK 6243

Qy      96 FTCSIQNI------SFSSFTLQRAGP 115
           :| :  |       : |   |:: ||
Db    6244 YTITATNSAGTATENLSVIVLEKPGP 6269


Search completed: November 14, 2002, 17:31:48
Job time : 24 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 14, 2002, 16:48:52 ; Search time 13 Seconds
(without alignments)
376.477 Million cell updates/sec

Title: US-09-598-443-2_COPY_1_118
Perfect score: 626
Sequence: 1 MPGVCDRAPDFLSPSEDQVL.....SIQNISFSSFTLQRAGPTSH 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	90.5	14.5	1709	1	SN_HUMAN	Q9bzz2 homo sapien
2	83.5	13.3	216	1	HIS5_BRUME	Q8ye35 brucella me
3	83.5	13.3	1694	1	SN_MOUSE	Q62230 mus musculu
4	82	13.1	628	1	LU_HUMAN	P50895 homo sapien
5	80.5	12.9	808	1	FGR4_MOUSE	Q03142 mus musculu
6	80	12.8	2012	1	DSCA_HUMAN	O60469 homo sapien
7	79.5	12.7	1242	1	NPHN_MOUSE	Q9qzs7 mus musculu
8	78.5	12.5	1234	1	NPHN_RAT	Q9r044 rattus norv
9	76.5	12.2	802	1	FGR4_HUMAN	P22455 homo sapien
10	76.5	12.2	1142	1	MYPF_HUMAN	Q14324 homo sapien
11	76.5	12.2	1241	1	NPHN_HUMAN	O60500 homo sapien
12	76	12.1	602	1	PGH1_MOUSE	P22437 mus musculu
13	75.5	12.1	822	1	FGR1_HUMAN	P11362 homo sapien
14	75.5	12.1	822	1	FGR1_MOUSE	P16092 mus musculu
15	75	12.0	359	1	LACH_DROME	Q24372 drosophila
16	74.5	11.9	819	1	FGR1_CHICK	P21804 gallus gall
17	74	11.8	824	1	MLT1_HUMAN	Q9udy8 homo sapien
18	73.5	11.7	1666	1	MYM1_MOUSE	Q62234 mus musculu
19	73	11.7	1447	1	DCC_MOUSE	P70211 mus musculu
20	73	11.7	3038	1	TRIO_HUMAN	O75962 homo sapien
21	72.5	11.6	806	1	CEK2_CHICK	P18460 gallus gall
22	72	11.5	248	1	MYP0_HUMAN	P25189 homo sapien
23	72	11.5	333	1	AMAL_DROME	P15364 drosophila
24	72	11.5	873	1	FAS2_DROME	P34082 drosophila
25	72	11.5	987	1	K6P1_CANAL	O94201 candida alb
26	72	11.5	1070	1	PTK7_HUMAN	Q13308 homo sapien
27	71.5	11.4	602	1	PGH1_RAT	Q63921 rattus norv
28	71	11.3	654	1	BFR2_HUMAN	Q01742 homo sapien
29	70.5	11.3	847	1	CD22_HUMAN	P20273 homo sapien
30	70.5	11.3	887	1	UFO_HUMAN	P30530 homo sapien
31	70.5	11.3	1897	1	PTPF_HUMAN	P10586 homo sapien
32	70	11.2	365	1	CXAR_HUMAN	P78310 homo sapien
33	70	11.2	1447	1	DCC_HUMAN	P43146 homo sapien
34	69.5	11.1	327	1	XYNC_EMENI	Q00177 emericella
35	69.5	11.1	402	1	RAGE_RAT	Q63495 rattus norv
36	69.5	11.1	498	1	VGLX_PRVRI	P07562 pseudorabie
37	69.5	11.1	599	1	PGH1_HUMAN	P23219 homo sapien
38	69.5	11.1	888	1	UFO_MOUSE	Q00993 mus musculu
39	69	11.0	248	1	MYP0_MOUSE	P27573 mus musculu
40	69	11.0	1010	1	SNI2_YEAST	P38163 saccharomyc
41	68.5	10.9	322	1	LDHB_FUNPA	P42122 fundulus pa
42	68.5	10.9	801	1	FGR3_MOUSE	Q61851 mus musculu
43	68.5	10.9	806	1	FGR3_HUMAN	P22607 homo sapien
44	68.5	10.9	813	1	FGR2_XENLA	Q03364 xenopus lae
45	68.5	10.9	822	1	FGR1_RAT	Q04589 rattus norv

ALIGNMENTS

RESULT 1
SN_HUMAN
ID SN_HUMAN STANDARD; PRT; 1709 AA.
AC Q9BZZ2; Q9H1H6; Q9H1H7; Q9H7L7; Q9GZS5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-
DE 1) (CD169 antigen).
GN SN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Hartnell A., Steel J., Turley H., Jones M., Jackson D., Crocker P.R.;
RT "Characterisation of human sialoadhesin (siglec-1), a sialic acid
RT binding receptor expressed on resident and inflammatory macrophage
RT populations.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehvaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 733-1709 FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Spleen;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human

```
RT   spleen.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MACROPHAGE-RESTRICTED ADHESION MOLECULE THAT MEDIATES
CC       SIALIC-ACID DEPENDENT BINDING TO CELLS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ISOFORM 1) AND
CC       SOLUBLE (ISOFORM 2).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=PROW 2:18-22(2001);
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/985165905_g.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF230073; AAK00757.1; -.
DR   EMBL; AL109804; CAC17543.1; -.
DR   EMBL; AL109804; CAC17542.1; -.
DR   EMBL; AK024462; BAB15752.1; -.
DR   EMBL; AK024459; BAB15749.1; -.
DR   EMBL; AK024479; BAB15769.1; -.
DR   HSSP; Q62230; 1QFO.
DR   Genew; HGNC:11127; SN.
DR   MIM; 600751; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 14.
DR   SMART; SM00409; IG; 17.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 14.
KW   Signal; Glycoprotein; Cell adhesion; Immunoglobulin domain; Repeat;
KW   Transmembrane; Alternative splicing; Antigen.
FT   SIGNAL        1     19       BY SIMILARITY.
FT   CHAIN        20   1709       SIALOADHESIN.
FT   DOMAIN       20   1641       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1642   1662       POTENTIAL.
FT   DOMAIN     1663   1709       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20    136       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      153    224       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      255    312       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      339    397       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      426    498       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      524    582       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      617    696       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      722    781       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      810    883       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      909    967       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      998   1074       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1100   1156       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN     1186   1248       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN     1274   1331       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN     1360   1432       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN     1458   1518       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN     1547   1620       IG-LIKE C2-TYPE DOMAIN 16.
FT   DISULFID     41     98       BY SIMILARITY.
FT   DISULFID    160    217       BY SIMILARITY.
FT   DISULFID    262    305       BY SIMILARITY.
FT   DISULFID    346    390       BY SIMILARITY.
FT   DISULFID    433    491       BY SIMILARITY.
FT   DISULFID    531    575       BY SIMILARITY.
FT   DISULFID    624    689       BY SIMILARITY.
FT   DISULFID    729    774       BY SIMILARITY.
FT   DISULFID    817    876       BY SIMILARITY.
FT   DISULFID    916    960       BY SIMILARITY.
FT   DISULFID   1005   1067       BY SIMILARITY.
FT   DISULFID   1107   1149       BY SIMILARITY.
FT   DISULFID   1193   1241       BY SIMILARITY.
FT   DISULFID   1281   1324       BY SIMILARITY.
FT   DISULFID   1367   1425       BY SIMILARITY.
FT   DISULFID   1465   1511       BY SIMILARITY.
FT   DISULFID   1554   1613       BY SIMILARITY.
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    265    265       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    697    697       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    726    726       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    730    730       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    741    741       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1104   1104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1138   1138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1251   1251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1462   1462       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1476   1476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1632   1709       ALHRLHQFQQLLWVLGLLVGLLLLLGLGACYTWRRRRVCK
FT                                QSMGENSVEMAFQKETTQLIDPDAATCETSTCAPPLG ->
FT                                GEGRGLHLPGHSAQKPSS (IN ISOFORM 2).
FT   CONFLICT   1349   1349       A -> T (IN REF. 1).
FT   CONFLICT   1519   1519       A -> V (IN REF. 3; BAB15749/BAB15769).
SQ   SEQUENCE   1709 AA;  182624 MW;  587C7CCA0B789A6D CRC64;

  Query Match             14.5%;  Score 90.5;  DB 1;  Length 1709;
  Best Local Similarity   22.8%;  Pred. No. 0.41;
  Matches   39;  Conservative   19;  Mismatches   38;  Indels   75;  Gaps    7;

Qy      12 LSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPL--------------- 56
           ||||   :|    |  | | |    |:  : :: |::|||||: |
Db     245 LSPSGRNILP---GELVTLTCQ---VNSSYPAVSSIKWLKDGVRLQTKTGVLHLPQAAWS 298

Qy      57 -----------GIGG----HYSLH----------------------------------EY 67
                      |:|       |||      .                            |
Db     299 DAGVYTCQAENGVGSLVSPPISLHIFMAEVQVSPAGPILENQTVTLVCNTPNEAPSDLRY 358

Qy      68 SWVKAN-LSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTS 117
           || | : | |   |  | ::: :    | : | :||:  |    :|:|| |
Db     359 SWYKNHVLLEDAHSHTLRLHLATRADTGFYFCEVQNVHGS----ERSGPVS 405


RESULT 2
HIS5_BRUME
ID   HIS5_BRUME     STANDARD;      PRT;   216 AA.
AC   Q8YE35;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Imidazole glycerol phosphate synthase subunit hisH (EC 2.4.2.-) (IGP
DE   synthase glutamine amidotransferase subunit) (IGP synthase subunit
DE   hisH) (ImGP synthase subunit hisH) (IGPS subunit hisH).
GN   HISH OR BMEI2043.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / Biotype 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC   -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
```

```
CC       IGP, AICAR and glutamate. The hisH subunit provides the glutamine
CC       amidotransferase activity that produces the ammonia necessary to
CC       hisF for the synthesis of IGP and AICAR (By similarity).
CC   -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
CC       ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC       carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC       aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC   -!- PATHWAY: Histidine biosynthesis; fifth step.
CC   -!- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009637; AAL53224.1; ALT_INIT.
DR   InterPro; IPR000991; GATase_1.
DR   PROSITE; PS00442; GATASE_TYPE_I; 1.
KW   Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW   Complete proteome.
FT   ACT_SITE     88     88       BY SIMILARITY.
FT   ACT_SITE    196    196       BY SIMILARITY.
FT   ACT_SITE    198    198       BY SIMILARITY.
SQ   SEQUENCE   216 AA;  23293 MW;  D7156BF1985C73CF CRC64;

  Query Match             13.3%;  Score 83.5;  DB 1;  Length 216;
  Best Local Similarity   24.6%;  Pred. No. 0.2;
  Matches   33;  Conservative   13;  Mismatches   43;  Indels   45;  Gaps    6;

Qy       1 MPGV---------CDRAPDFLSPSEDQVL---RPALGSSVAL-------------NCTAW 35
           :|||         |  |   :   | ||   || ||  | :             |  |
Db      50 LPGVGAYADCRRGLDAVPGMVEALNDTVLKKARPFLGICVGMQLMSERGLEKTVTNGLGW 109

Qy      36 VV--------SGPHCSLPSVQW----------LKDGLPLGIGG--HYSLHEYSWVKANLS 75
           :         |     :| : |          : ||:| |  |   | :| |     | |
Db     110 IAGDVREMVPSDASLKIPQIGWNRIHVKHSHPIFDGIPTGDDGLHAYFVHSYMLDAKNAS 169

Qy      76 EVLVSSVLGVNVTS 89
           :||  :  | :||:
Db     170 DVLAVTDYGGDVTA 183


ULT 3
_MOUSE
ID   SN_MOUSE       STANDARD;      PRT;  1694 AA.
AC   Q62230; O55216; Q62228; Q62229;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-
DE   1) (Sheep erythrocyte receptor) (SER).
GN   SN OR SA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=C57BL/6; TISSUE=Macrophage;
RX   MEDLINE=95009950; PubMed=7925291;
RA   Crocker P.R., Mucklow S., Boukson V., McWilliam A., Willis A.C.,
RA   Gordon S., Milon G., Kelm S., Bradfield P.;
RT   "Sialoadhesin, a macrophage sialic acid binding receptor for
RT   haemopoietic cells with 17 immunoglobulin-like domains.";
RL   EMBO J. 13:4490-4503(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98051930; PubMed=9383289;
RA   Mucklow S., Gordon S., Crocker P.R.;
RT   "Characterization of the mouse sialoadhesin gene, Sn.";
RL   Mamm. Genome 8:934-937(1997).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=C57BL/6; TISSUE=Spleen;
RX   MEDLINE=91266893; PubMed=2050106;
RA   Crocker P.R.;
RT   "Purification and properties of sialoadhesin, a sialic acid-binding
RT   receptor of murine tissue macrophages.";
RL   EMBO J. 10:1661-1669(1991).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.
RX   MEDLINE=98325385; PubMed=9660955;
RA   May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;
RT   "Crystal structure of the N-terminal domain of sialoadhesin in
RT   complex with 3' sialyllactose at 1.85 A resolution.";
RL   Mol. Cell 1:719-728(1998).
CC   -!- FUNCTION: MACROPHAGE-RESTRICTED ADHESION MOLECULE THAT MEDIATES
CC       SIALIC-ACID DEPENDENT BINDING TO CELLS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ISOFORM 1) AND
CC       SOLUBLE (ISOFORMS 2 AND 3).
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SPLEEN AND LYMPH NODE
CC       WITH LOWER AMOUNTS IN LUNG, LIVER, BONE MARROW, HEART AND SKIN. NO
CC       EXPRESSION IN THYMUS, KIDNEY, BRAIN OR SMALL INTESTINE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36293; CAA85290.1; -.
DR   EMBL; Z36233; CAA85268.1; -.
DR   EMBL; Z36234; CAA85269.1; -.
DR   EMBL; U92842; AAB95641.1; -.
DR   EMBL; U92833; AAB95641.1; JOINED.
DR   EMBL; U92834; AAB95641.1; JOINED.
DR   EMBL; U92836; AAB95641.1; JOINED.
DR   EMBL; U92837; AAB95641.1; JOINED.
DR   EMBL; U92838; AAB95641.1; JOINED.
DR   EMBL; U92839; AAB95641.1; JOINED.
DR   EMBL; U92840; AAB95641.1; JOINED.
DR   EMBL; U92841; AAB95641.1; JOINED.
DR   PDB; 1QFO; 16-APR-99.
DR   PDB; 1QFP; 16-APR-99.
DR   MGD; MGI:99668; Sn.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 16.
DR   SMART; SM00410; IG_like; 9.
DR   SMART; SM00408; IGc2; 6.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Signal; Glycoprotein; Cell adhesion; Immunoglobulin domain; Repeat;
KW   Transmembrane; Alternative splicing; 3D-structure.
FT   SIGNAL        1     19
FT   CHAIN        20   1694       SIALOADHESIN.
FT   DOMAIN       20   1638       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1639   1659       POTENTIAL.
FT   DOMAIN     1660   1694       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20    136       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      153    225       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      256    313       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      340    398       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      427    499       IG-LIKE C2-TYPE DOMAIN 4.
```

```
FT   DOMAIN      525    583       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      618    692       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      718    777       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      806    879       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      904    962       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      993   1069       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1095   1151       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN     1181   1243       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN     1269   1326       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN     1355   1428       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN     1455   1515       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN     1544   1617       IG-LIKE C2-TYPE DOMAIN 16.
FT   SITE        827    829       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     41     98
FT   DISULFID    160    218       BY SIMILARITY.
FT   DISULFID    263    306       BY SIMILARITY.
FT   DISULFID    347    391       BY SIMILARITY.
FT   DISULFID    434    492       BY SIMILARITY.
FT   DISULFID    532    576       BY SIMILARITY.
FT   DISULFID    625    685       BY SIMILARITY.
FT   DISULFID    725    770       BY SIMILARITY.
FT   DISULFID    813    872       BY SIMILARITY.
FT   DISULFID    911    955       BY SIMILARITY.
FT   DISULFID   1000   1062       BY SIMILARITY.
FT   DISULFID   1102   1144       BY SIMILARITY.
FT   DISULFID   1188   1236       BY SIMILARITY.
FT   DISULFID   1276   1319       BY SIMILARITY.
FT   DISULFID   1362   1421       BY SIMILARITY.
FT   DISULFID   1462   1508       BY SIMILARITY.
FT   DISULFID   1551   1610       BY SIMILARITY.
FT   CARBOHYD    159    159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    500    500       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    583    583       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    693    693       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    722    722       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    882    882       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1099   1099       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1459   1459       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1473   1473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    326    340       MAEVKMNPAGPVLEN -> SESWMRLRGPVSGKH (IN
FT                                ISOFORM 2).
FT   VARSPLIC    341   1694       MISSING (IN ISOFORM 2).
FT   VARSPLIC   1528   1598       YPPKTPTLIVFVEPQGGHQGILDCRVDSEPLAILTLHRGSQ
                                  LVASNQLHDAPTKPHIRVTAPPNALRVDIE -> CEYEPIS
                                  ALCLSLHLTGPYQAFSSAQSKGFIGKGLRTLASSLAGCMWF
FT                                VSMLGYPALKWRILLPFWDEYRR (IN ISOFORM 3).
FT   VARSPLIC   1599   1694       MISSING (IN ISOFORM 3).
FT   CONFLICT    590    590       P -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT   1049   1051       IHF -> FLV (IN REF. 1; AA SEQUENCE).
FT   CONFLICT   1054   1055       LE -> VQ (IN REF. 1; AA SEQUENCE).
FT   CONFLICT   1061   1061       T -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT   1065   1065       S -> Q (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   1694 AA;  183087 MW;  1D2BFB39B7299CBC CRC64;

  Query Match             13.3%;  Score 83.5;  DB 1;  Length 1694;
  Best Local Similarity   21.6%;  Pred. No. 2.1;
  Matches   37;  Conservative   17;  Mismatches   42;  Indels   75;  Gaps    6;

Qy      12 LSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGH---------- 61
           || |   :|    |  | | |    |:  : :: :||| :||: ||: ||
Db     246 LSSSGRNILP---GDPVTLTCR---VNSSYPAVSAVQWARDGVNLGVTGHVLRLFSAAWN 299

Qy      62 --------------------YSLH---------------------------------EY 67
                              |||                                 |
Db     300 DSGAYTCQATNDMGSLVSSPLSLHVFMAEVKMNPAGPVLENETVTLLCSTPKEAPQELRY 359

Qy      68 SWVKAN-LSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTS 117
```

```
           || | : | |   :| | :   :    | : | :||   |    :|: | |
Db     360 SWYKNHILLEDAHASTLHLPAVTRADTGFYFCEVQNAQGS----ERSSPLS 406


RESULT 4
LU_HUMAN
ID   LU_HUMAN       STANDARD;      PRT;   628 AA.
AC   P50895;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lutheran blood group glycoprotein precursor (B-CAM cell surface
DE   glycoprotein) (Auberger B antigen) (F8/G253 antigen).
GN   LU OR BCAM OR MSK19.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
RC   TISSUE=Placenta;
RX   MEDLINE=95296337; PubMed=7777537;
RA   Parsons S.F., Mallinson G., Holmes C.H., Houlihan J.M., Simpson K.L.,
RA   Mawby W.J., Spurr N.K., Warne D., Barclay A.N., Anstee D.J.;
RT   "The Lutheran blood group glycoprotein, another member of the
RT   immunoglobulin superfamily, is widely expressed in human tissues and
RT   is developmentally regulated in human liver.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
RN   [2]
RP   SEQUENCE OF 1-588 FROM N.A.
RX   MEDLINE=95042297; PubMed=7954395;
RA   Campbell I.G., Foulkes W.D., Senger G., Trowsdale J.,
RA   Garin-Chesa P., Rettig W.J.;
RT   "Molecular cloning of the B-CAM cell surface glycoprotein of
RT   epithelial cancers: a novel member of the immunoglobulin
RT   superfamily.";
RL   Cancer Res. 54:5761-5765(1994).
CC   -!- FUNCTION: PROBABLE RECEPTOR. MAY MEDIATE INTRAELLULAR SIGNALING.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION (HIGHEST IN THE
CC       PANCREAS AND VERY LOW IN BRAIN). CLOSELY ASSOCIATED WITH THE BASAL
CC       LAYER OF CELLS IN EPITHELIA AND THE ENDOTHELIUM OF BLOOD VESSEL
CC       WALLS.
CC   -!- DEVELOPMENTAL STAGE: IS UNDER DEVELOPMENTAL CONTROL IN LIVER AND
CC       MAY ALSO BE REGULATED DURING DIFFERENTIATION IN OTHER TISSUES.
CC       UPREGULATED FOLLOWING MALIGNANT TRANSFORMATION IN SOME CELL TYPES.
CC   -!- POLYMORPHISM: LU IS RESPONSIBLE FOR THE LUTHERAN BLOOD GROUP
CC       SYSTEM.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83425; CAA58449.1; -.
DR   EMBL; X80026; CAA56327.1; -.
DR   Genew; HGNC:6722; LU.
DR   MIM; 111200; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 2.
KW   Receptor; Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;
KW   Repeat; Blood group antigen.
FT   SIGNAL        1     31
```

```
FT   CHAIN        32    628       LUTHERAN BLOOD GROUP GLYCOPROTEIN.
FT   DOMAIN       32    547       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    548    568       POTENTIAL.
FT   DOMAIN       569    628       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        46    132       IG-LIKE V-TYPE DOMAIN 1.
FT   DOMAIN       165    244       IG-LIKE V-TYPE DOMAIN 2.
FT   DOMAIN       284    344       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       377    431       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       466    529       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID      53    125       PROBABLE.
FT   DISULFID     172    237       PROBABLE.
FT   DISULFID     291    337       PROBABLE.
FT   DISULFID     384    424       PROBABLE.
FT   DISULFID     473    522       PROBABLE.
FT   CARBOHYD     321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     383    383       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     439    439       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     225    226       RL -> PC (IN REF. 2).
FT   CONFLICT     355    356       EL -> DV (IN REF. 2).
SQ   SEQUENCE   628 AA;  67374 MW;  C89B0A4835492B1E CRC64;

  Query Match             13.1%;  Score 82;  DB 1;  Length 628;
  Best Local Similarity   36.2%;  Pred. No. 0.95;
  Matches   29;  Conservative   12;  Mismatches   33;  Indels    6;  Gaps    3;

Qy      10 DFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSW 69
           | |  || :||   | ||  :||:   | |     |:::| ||  ||| |   ||   ::
Db     362 DPLELSEGKVLSLPLNSSAVVNCS---VHG--LPTPALRWTKDSTPLGDGPMLSLSSITF 416

Qy      70 VKANLSEVLVSSVLGVNVTS 89
             :| : |  :|:  | | |
Db     417 -DSNGTYVCEASLPTVPVLS 435


RESULT 5
FGR4_MOUSE
ID   FGR4_MOUSE     STANDARD;      PRT;   808 AA.
AC   Q03142;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fibroblast growth factor receptor 4 precursor (EC 2.7.1.112) (FGFR-4)
DE   (Protein-tyrosine kinase receptor MPK-11).
GN   FGFR4 OR FGFR-4 OR MPK-11.
     Mus musculus (Mouse).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Fetal cerebellum;
RX   MEDLINE=92146274; PubMed=1723680;
RA   Stark K.L., McMahon J., McMahon A.P.;
RT   "FGFR-4, a new member of the fibroblast growth factor receptor
RT   family, expressed in the definitive endoderm and skeletal muscle
RT   lineages of the mouse.";
RL   Development 113:641-651(1991).
RN   [2]
RP   SEQUENCE OF 620-676 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Embryonic brain;
RX   MEDLINE=93096484; PubMed=1281307;
RA   Gilardi-Hebenstreit P., Nieto M.A., Frain M., Mattei M.-G.,
RA   Chestier A., Wilkinson D.G., Charnay P.;
RT   "An Eph-related receptor protein tyrosine kinase gene segmentally
RT   expressed in the developing mouse hindbrain.";
RL   Oncogene 7:2499-2506(1992).
CC   -!- FUNCTION: PUTATIVE RECEPTOR FOR BASIC FIBROBLAST GROWTH FACTOR.
CC       MAY BE INVOLVED IN THE DEVELOPMENT OF SKELETAL MUSCLE CELL
CC       LINEAGES.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING GUT ENDODERM,
CC       IN MYOTOMALLY DERIVED SKELETAL MUSCLE, THE ADRENAL CORTEX,
CC       KIDNEY AND CONDENSING CARTILAGE. PRESENT IN ADULT LIVER,
CC       LUNG AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59927; CAA42551.1; -.
DR   EMBL; X57236; CAA40512.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95525; Fgfr4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW   Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW   Repeat; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    808       FIBROBLAST GROWTH FACTOR RECEPTOR 4.
FT   DOMAIN       19    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    367    387       POTENTIAL.
FT   DOMAIN      388    808       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       53     98       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      169    221       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      268    330       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      464    761       PROTEIN KINASE.
FT   NP_BIND     470    478       ATP (BY SIMILARITY).
FT   BINDING     509    509       ATP (BY SIMILARITY).
FT   ACT_SITE    618    618       BY SIMILARITY.
FT   MOD_RES     649    649       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   DISULFID     53     98       POTENTIAL.
FT   DISULFID    169    221       POTENTIAL.
FT   DISULFID    268    330       POTENTIAL.
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    255    255       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    308    308       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   808 AA;  89775 MW;  E3F30BE5E334E8E6 CRC64;

  Query Match             12.9%;  Score 80.5;  DB 1;  Length 808;
  Best Local Similarity   30.8%;  Pred. No. 1.8;
  Matches   28;  Conservative   11;  Mismatches   37;  Indels   15;  Gaps    4;

Qy      24 LGSSVALNCTAWVVSGPHCSLPSVQWLK----DGLPLGIGGHYSLHEYSWVKANLSEVLV 79
           :|| | | |  :  : ||     :||||    :|   |  |   :        |:|||
Db     260 VGSDVELLCKVYSDAQPH-----IQWLKHVVINGSSFGADGFPYVQVLKTTDINISEV-- 312

Qy      80 SSVLGVNVTSTEVYGAFTCSIQN---ISFSS 107
             || :   | |  | :||   |   :|: |
Db     313 -QVLYLRNVSAEDAGEYTCLAGNSIGLSYQS 342
```

```
RESULT 6
DSCA_HUMAN
ID   DSCA_HUMAN     STANDARD;      PRT;  2012 AA.
AC   O60469; O60468;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Down syndrome cell adhesion molecule precursor (CHD2).
GN   DSCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=98087574; PubMed=9426258;
RA   Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA   Lyons G.E., Korenberg J.R.;
RT   "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT   Down syndrome region and is involved in the development of the
[redacted]
RT   nervous system.";
RL   Hum. Mol. Genet. 7:227-237(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=20384934; PubMed=10925149;
RA   Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT   "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT   intercellular adhesion.";
RL   Brain Res. Mol. Brain Res. 79:118-126(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC       INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
[redacted]
CC       NERVOUS SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC       SHORT ISOFORM MAY BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC       AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023450; AAC17967.1; -.
DR   EMBL; AF023449; AAC17966.1; -.
DR   EMBL; AF217525; AAF27525.1; -.
DR   EMBL; AL163283; CAB90464.1; -.
DR   EMBL; AL163282; CAB90436.1; -.
DR   EMBL; AL163281; CAB90444.1; -.
DR   Genew; HGNC:3039; DSCAM.
DR   MIM; 602523; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 7.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2012       DOWN SYNDROME CELL ADHESION MOLECULE.
FT   DOMAIN       18   1595       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   DOMAIN     1617   2012       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    109       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      138    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      328    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      421    491       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    582       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      610    676       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      704    773       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      802    872       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      885    972       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      984   1076       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     1088   1177       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1189   1273       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1300   1366       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1380   1463       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1477   1562       FIBRONECTIN TYPE-III 6.
FT   DISULFID     46    102       BY SIMILARITY.
FT   DISULFID    145    197       BY SIMILARITY.
FT   DISULFID    246    293       BY SIMILARITY.
FT   DISULFID    335    385       BY SIMILARITY.
FT   DISULFID    428    484       BY SIMILARITY.
FT   DISULFID    525    575       BY SIMILARITY.
FT   DISULFID    617    669       BY SIMILARITY.
FT   DISULFID    711    766       BY SIMILARITY.
FT   DISULFID    809    865       BY SIMILARITY.
FT   DISULFID   1307   1359       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    512    512       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    748    748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    795    795       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1160   1160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1250   1250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1271   1271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1488   1488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1562   1571       NFATLNYDGS -> KEAARCKEFS (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1572   2012       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1893   2012       HRPGDLIHLPPYLRMDFLLNRGGPGTSRDLSLGQACLEPQK
FT                                SRTLKRPTVLEPIPMEAASSASSTREGQSWQPGAVATLPQR
FT                                EGAELGQAAKMSSSQESLLDSRGHLKGNNPYAKSYTLV ->
FT                                IGQVTSYICLHTLEWTFC (IN REF. 1).
SQ   SEQUENCE   2012 AA;  222259 MW;  0E33CFB781A08334 CRC64;
```

```
  Query Match           12.8%;  Score 80;  DB 1;  Length 2012;
  Best Local Similarity  26.3%;  Pred. No. 5.6;
  Matches   26;  Conservative   19;  Mismatches   38;  Indels   16;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLH 65
           | |  :|   ::|: ||    |:| |    | |    ||:: |  |  |:  || : :
Db     404 DGTPKIISAFSEKVVSPA--EPVSLMCN---VKG--TPLPTITWTLDDDPILKGGSHRIS 456

Qy      66 EYSWVKANLSEVLVSSVLGVNVTSTEVY--GAFTCSIQN 102
           :    : |:   |       |::|::|   | : |:  |
Db     457 QMITSEGNVVSYL-------NISSSQVRDGGVYRCTANN 488


RESULT 7
NPHN_MOUSE
ID   NPHN_MOUSE     STANDARD;      PRT;  1242 AA.
AC   Q9QZS7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nephrin precursor (Renal glomerulus-specific cell adhesion receptor).
GN   NPHS1 OR NPHN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RX   MEDLINE=99436348; PubMed=10504499;
RA   Holzman L.B., St John P.L., Kovari I.A., Verma R., Holthoefer H.,
RA   Abrahamson D.R.;
RT   "Nephrin localizes to the slit pore of the glomerular epithelial
RT   cell.";
RL   Kidney Int. 56:1481-1491(1999).
RN   [2]
RP   INTERACTION WITH CD2AP.
RX   MEDLINE=21590051; PubMed=11733379;
RA   Shih N.Y., Li J., Cotran R., Mundel P., Miner J.H., Shaw A.S.;
RT   "CD2AP localizes to the slit diaphragm and binds to nephrin via a
RT   novel C-terminal domain.";
RL   Am. J. Pathol. 159:2303-2308(2001).
RN   [3]
RP   INTERACTION WITH CD2AP AND NPHS2.
RX   MEDLINE=21590460; PubMed=11733557;
RA   Schwarz K., Simons M., Reiser J., Saleem M.A., Faul C., Kriz W.,
RA   Shaw A.S., Holzman L.B., Mundel P.;
RT   "Podocin, a raft-associated component of the glomerular slit
RT   diaphragm, interacts with CD2AP and nephrin.";
RL   J. Clin. Invest. 108:1621-1629(2001).
CC   -!- FUNCTION: Seems to play a role in the development or function of
CC       the kidney glomerular filtration barrier. May anchor the podocyte
CC       slit diaphragm to the actin cytoskeleton.
CC   -!- SUBUNIT: Interacts with podocin/NPHS2 and with CD2AP C-terminal
CC       domain.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). Located
CC       at podocyte slit diaphragm between podocyte foot processes.
CC   -!- TISSUE SPECIFICITY: Expressed in kidney glomeruli.
CC   -!- PTM: Phosphorylated on tyrosine residues (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 8 IMMUNOGLOBULIN-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF168466; AAF03368.1; -.
DR   MGD; MGI:1859637; Nphs1.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR000399; TPP_enzyme.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 3.
KW   Cell adhesion; Transmembrane; Signal; Glycoprotein;
KW   Immunoglobulin domain; Repeat; Phosphorylation.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1242       NEPHRIN.
FT   DOMAIN       23   1064       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1065   1086       POTENTIAL.
FT   DOMAIN     1087   1242       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       46    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      153    224       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      258    324       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      354    424       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      458    535       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      560    630       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      754    823       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      856    927       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      941   1025       FIBRONECTIN TYPE-III.
FT   DISULFID     53    111       POTENTIAL.
FT   DISULFID    160    217       POTENTIAL.
FT   DISULFID    265    317       POTENTIAL.
FT   DISULFID    361    417       POTENTIAL.
FT   DISULFID    465    528       POTENTIAL.
FT   DISULFID    567    623       POTENTIAL.
FT   DISULFID    761    816       POTENTIAL.
FT   DISULFID    863    920       POTENTIAL.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    577    577       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    908    908       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1242 AA;  134890 MW;  02D82180BF145092 CRC64;

  Query Match           12.7%;  Score 79.5;  DB 1;  Length 1242;
  Best Local Similarity  29.5%;  Pred. No. 3.7;
  Matches   26;  Conservative   16;  Mismatches   29;  Indels   17;  Gaps    5;

Qy      26 SSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGI-GGHYSLHEYSWVKANLSEVLVSSV 82
           ||  |:| |         :|::   | |:|:|| :    |: |:|         |: ||:
Db     857 SSATLHCRA-------RGVPNIDFTWTKNGVPLDLQDPRYTEHKYH------QGVVHSSL 903

Qy      83 LGV-NVTSTEVYGAFTCSIQNISFSSFT 109
           | : ||:: : |  | |:  |   |  |
Db     904 LTIANVSAAQDYALFKCTATNALGSDHT 931


RESULT 8
NPHN_RAT
ID   NPHN_RAT       STANDARD;      PRT;  1234 AA.
AC   Q9R044; Q9QXX7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nephrin precursor (Renal glomerulus-specific cell adhesion receptor).
GN   NPHS1 OR NPHN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
```

```
RC   STRAIN=Sprague-Dawley; TISSUE=Renal glomerulus;
RX   MEDLINE=99419288; PubMed=10487848;
RA   Ahola H., Wang S.-X., Luimula P., Solin M.-L., Holzman L.B.,
RA   Holthoefer H.;
RT   "Cloning and expression of the rat nephrin homolog.";
RL   Am. J. Pathol. 155:907-913(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.
RC   STRAIN=Wistar; TISSUE=Renal glomerulus;
RX   MEDLINE=20253275; PubMed=10792613;
RA   Kawachi H., Koike H., Kurihara H., Yaoita E., Orikasa M., Shia M.A.,
RA   Sakai T., Yamamoto T., Salant D.J., Shimizu F.;
RT   "Cloning of rat nephrin: expression in developing glomeruli and in
RT   proteinuric states.";
RL   Kidney Int. 57:1949-1961(2000).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=21868269; PubMed=11880318;
RA   Yuan H., Takeuchi E., Salant D.J.;
RT   "Podocyte slit-diaphragm protein nephrin is linked to the actin
RT   cytoskeleton.";
RL   Am. J. Physiol. 282:F585-F591(2002).
CC   -!- FUNCTION: Seems to play a role in the development or function of
CC       the kidney glomerular filtration barrier. May anchor the podocyte
CC       slit diaphragm to the cytoskeleton.
CC   -!- SUBUNIT: Interacts with podocin/NPHS2 and with CD2AP C-terminal
CC       domain (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). Located
CC       at podocyte slit diaphragm between podocyte foot processes.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; isoform 1 (shown here), 2/alpha
CC       and 3/beta; may be produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Specifically expressed in the podocytes of
CC       kidney glomeruli.
CC   -!- PTM: Phosphorylated on tyrosine residues (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 8 IMMUNOGLOBULIN-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125521; AAF12734.1; -.
DR   EMBL; AF161715; AAF14884.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 6.
KW   Cell adhesion; Transmembrane; Signal; Glycoprotein;
KW   Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1234       NEPHRIN.
FT   DOMAIN       19   1060       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1061   1081       POTENTIAL.
FT   DOMAIN     1082   1234       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       42    114       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      149    220       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    320       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      350    420       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      454    531       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      556    626       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      750    819       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      852    923       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      938   1022       FIBRONECTIN TYPE-III.
FT   DISULFID     49    107       POTENTIAL.
FT   DISULFID    156    213       POTENTIAL.
FT   DISULFID    261    313       POTENTIAL.
FT   DISULFID    357    413       POTENTIAL.
FT   DISULFID    461    524       POTENTIAL.
FT   DISULFID    563    619       POTENTIAL.
FT   DISULFID    757    812       POTENTIAL.
FT   DISULFID    859    916       POTENTIAL.
FT   CARBOHYD     36     36       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    352    352       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    397    397       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    549    549       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    704    704       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    904    904       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1052   1064       MISSING (IN ISOFORM 2).
FT   VARSPLIC   1052   1137       MISSING (IN ISOFORM 3).
FT   CONFLICT      1      1       M -> MGAKRVTVRGARTSPIHRT (IN REF. 2).
FT   CONFLICT     28     28       R -> Q (IN REF. 2).
FT   CONFLICT   1211   1211       N -> D (IN REF. 2).
SQ   SEQUENCE   1234 AA;  134304 MW;  6BF6707A229CA51E CRC64;

  Query Match             12.5%;  Score 78.5;  DB 1;  Length 1234;
  Best Local Similarity   29.5%;  Pred. No. 4.6;
  Matches   26;  Conservative   15;  Mismatches   30;  Indels   17;  Gaps    5;

Qy      26 SSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGI-GGHYSLHEYSWVKANLSEVLVSSV 82
           ||  |:| |         :|::   | |:|:|| :    |: | |         |: ||:
Db     853 SSATLHCRA-------RGVPNIDFTWTKNGVPLDLQDPRYTEHRYH------QGVVHSSL 899

Qy      83 LGV-NVTSTEVYGAFTCSIQNISFSSFT 109
           | : ||:: : |  | |:  |   |  |
Db     900 LTIANVSAAQDYALFKCTATNALGSDHT 927


RESULT 9
FGR4_HUMAN
ID   FGR4_HUMAN     STANDARD;      PRT;   802 AA.
AC   P22455; Q14309; O43785;
DT   01-AUG-1991 (Rel. 19, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fibroblast growth factor receptor 4 precursor (EC 2.7.1.112) (FGFR-4).
GN   FGFR4 OR JTK2 OR TKF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91224085; PubMed=1709094;
RA   Partanen J.M., Maekelae T.P., Eerola E., Korhonen J., Hirvonen H.,
RA   Claesson-Welsh L., Alitalo K.;
RT   "FGFR-4, a novel acidic fibroblast growth factor receptor with a
RT   distinct expression pattern.";
RL   EMBO J. 10:1347-1354(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=93194827; PubMed=7680645;
RA   Ron D., Reich R., Chedid M., Lengel C., Cohen O.E., Chan A.M.,
RA   Neufeld G., Miki T., Tronick S.R.;
RT   "Fibroblast growth factor receptor 4 is a high affinity receptor for
RT   both acidic and basic fibroblast growth factor but not for
RT   keratinocyte growth factor.";
RL   J. Biol. Chem. 268:5388-5394(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98119018; PubMed=9457674;
RA   Kostrzewa M., Muller U.;
```

```
RT   "Genomic structure and complete sequence of the human FGFR4 gene.";
RL   Mamm. Genome 9:131-135(1998).
RN   [4]
RP   SEQUENCE OF 609-676 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=91062389; PubMed=2247464;
RA   Partanen J., Maekelae T.P., Alitalo R., Lehvaeslaiho H., Alitalo K.;
RT   "Putative tyrosine kinases expressed in K-562 human leukemia cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8913-8917(1990).
CC   -!- FUNCTION: RECEPTOR FOR ACIDIC FIBROBLAST GROWTH FACTOR. DOES NOT
CC       BIND TO BASIC FIBROBLAST GROWTH FACTOR. BINDS FGF19.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57205; CAA40490.1; -.
DR   EMBL; L03840; AAB59389.1; -.
DR   EMBL; Y13901; CAA74200.1; -.
DR   EMBL; M59373; AAA63208.1; -.
DR   PIR; S15345; TVHUF4.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:3691; FGFR4.
DR   MIM; 134935; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW   Transferase; Phosphorylation; Transmembrane; Immunoglobulin domain;
KW   Repeat; Signal.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    802       FIBROBLAST GROWTH FACTOR RECEPTOR 4.
FT   DOMAIN      25    369       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   370    390       POTENTIAL.
FT   DOMAIN     391    802       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      50    108       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     165    231       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     264    340       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN     467    755       PROTEIN KINASE.
FT   NP_BIND    473    481       ATP (BY SIMILARITY).
FT   BINDING    503    503       ATP (BY SIMILARITY).
FT   ACT_SITE   612    612       BY SIMILARITY.
FT   MOD_RES    643    643       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   DISULFID    57    101       POTENTIAL.
FT   DISULFID   172    224       POTENTIAL.
FT   DISULFID   271    333       POTENTIAL.
FT   CARBOHYD   112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   258    258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   311    311       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   322    322       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   297    297       D -> V (IN REF. 1).
SQ   SEQUENCE   802 AA;  87954 MW;  B22B259831BB889F CRC64;

  Query Match             12.2%;  Score 76.5;  DB 1;  Length 802;
  Best Local Similarity   30.8%;  Pred. No. 4.5;
  Matches   28;  Conservative   10;  Mismatches   38;  Indels   15;  Gaps    4;

Qy      24 LGSSVALNCTAWVVSGPHCSLPSVQWLK----DGLPLGIGGHYSLHEYSWVKANLSEVLV 79
           :|| | | |  :  : ||     :||||    :|   |  |   :         | |||
Db     263 VGSDVELLCKVYSDAQPH-----IQWLKHIVINGSSFGADGFPYVQVLKTADINSSEV-- 315

Qy      80 SSVLGVNVTSTEVYGAFTCSIQN---ISFSS 107
             || :   | |  | :||   |   :|: |
Db     316 -EVLYLRNVSAEDAGEYTCLAGNSIGLSYQS 345


RESULT 10
MYPF_HUMAN
ID   MYPF_HUMAN     STANDARD;      PRT;  1142 AA.
AC   Q14324;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin-binding protein C, fast-type (Fast MyBP-C) (C-protein, skeletal
DE   muscle fast-isoform).
GN   MYBPC2 OR MYBPCF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal skeletal muscle;
RX   MEDLINE=93387319; PubMed=8375400;
RA   Weber F.E., Vaughan K.T., Reinach F.C., Fischman D.A.;
RT   "Complete sequence of human fast-type and slow-type muscle myosin-
RT   binding-protein C (MyBP-C). Differential expression, conserved domain
RT   structure and chromosome assignment.";
RL   Eur. J. Biochem. 216:661-669(1993).
CC   -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
CC       CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
CC       IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
CC       ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
CC       CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. MYBP
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73113; CAA51544.1; -.
DR   HSSP; P56276; 1TLK.
DR   Genew; HGNC:7550; MYBPC2.
DR   MIM; 160793; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 3.
```

```
DR   SMART; SM00410; IG_like; 6.
DR   SMART; SM00408; IGc2; 1.
KW   Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
KW   Thick filament; Repeat.
FT   DOMAIN       50    153       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      255    344       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      345    437       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      438    539       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      540    639       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      640    737       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      738    838       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      839    933       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      934   1048       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1049   1142       IG-LIKE C2-TYPE DOMAIN 7.
SQ   SEQUENCE   1142 AA;  128142 MW;  93461D4356E86C09 CRC64;

  Query Match             12.2%;  Score 76.5;  DB 1;  Length 1142;
  Best Local Similarity   27.8%;  Pred. No. 6.7;
  Matches   27;  Conservative   13;  Mismatches   36;  Indels   21;  Gaps    5;

Qy       8 APDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEY 67
           || ||:|  |:|:    | | |||   |   |    | | |:|: : :    : :  |
Db    1048 APKFLTPLIDRVV--VAGYSAALNCA--VRGHPK---PKVVWMKNKMEIREDPKFLITNY 1100

Qy      68 SWVKANLSEVLVSSVLGVNVTSTEVY--GAFTCSIQN 102
                      || :|:     :  | :||   |
Db    1101 ------------QGVLTLNIRRPSPFDAGTYTCRAVN 1125


RESULT 11
NPHN_HUMAN
ID   NPHN_HUMAN     STANDARD;      PRT;  1241 AA.
AC   O60500;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nephrin precursor (Renal glomerulus-specific cell adhesion receptor).
GN   NPHS1 OR NPHN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=98325371; PubMed=9660941;
RA   Kestilae M., Lenkkeri U., Maennikkoe M., Lamerdin J., McCready P.,
RA   Putaala H., Ruotsalainen V., Morita T., Nissinen M., Herva R.,
RA   Kashtan C.E., Peltonen L., Holmberg C., Olsen A., Tryggvason K.;
RT   "Positionally cloned gene for a novel glomerular protein -- nephrin --
RT   is mutated in congenital nephrotic syndrome.";
RL   Mol. Cell 1:575-582(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Grunkemeyer J.A., Kumar N., Kalluri R.;
RT   "Human nephrin (NPHS1) cDNA sequence.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1032-1134 FROM N.A. (ISOFORM 2), AND SUBCELLULAR LOCATION.
RX   MEDLINE=20019662; PubMed=10550324;
RA   Holthoefer H., Ahola H., Solin M.-L., Wang S.-X., Palmen T.,
RA   Luimula P., Miettinen A., Kerjaschki D.;
RT   "Nephrin localizes at the podocyte filtration slit area and is
RT   characteristically spliced in the human kidney.";
RL   Am. J. Pathol. 155:1681-1687(1999).
RN   [4]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=99324171; PubMed=10393930;
RA   Ruotsalainen V., Ljungberg P., Wartiovaara J., Lenkkeri U.,
RA   Kestilae M., Jalanko H., Holmberg C., Tryggvason K.;
RT   "Nephrin is specifically located at the slit diaphragm of glomerular
RT   podocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:7962-7967(1999).
RN   [5]
RP   INTERACTION WITH NPHS2.
RX   MEDLINE=21551283; PubMed=11562357;
RA   Huber T.B., Kottgen M., Schilling B., Walz G., Benzing T.;
RT   "Interaction with podocin facilitates nephrin signaling.";
RL   J. Biol. Chem. 276:41543-41546(2001).
RN   [6]
RP   VARIANTS K-117 AND S-1077, AND VARIANTS CNF S-64; N-171; T-172 DEL;
RP   N-173; 205-T-P-R-207 DELINS I; C-270; P-350; R-366; C-367; S-368;
RP   V-376; Q-408; Y-465; F-528; Q-610; F-623; C-724; C-743; W-802; P-802;
RP   D-806; C-831 AND C-1140.
RX   MEDLINE=99115081; PubMed=9915943;
RA   Lenkkeri U., Maennikkoe M., McCready P., Lamerdin J., Gribouval O.,
RA   Niaudet P.M., Antignac C.K., Kashtan C.E., Homberg C., Olsen A.,
RA   Kestilae M., Tryggvason K.;
RT   "Structure of the gene for congenital nephrotic syndrome of the
RT   finnish type (NPHS1) and characterization of mutations.";
RL   Am. J. Hum. Genet. 64:51-61(1999).
RN   [7]
RP   VARIANTS CNF LYS-447 AND VAL-819.
RX   MEDLINE=20117947; PubMed=10652016;
RA   Aya K., Tanaka H., Seino Y.;
RT   "Novel mutation in the nephrin gene of a Japanese patient with
RT   congenital nephrotic syndrome of the Finnish type.";
RL   Kidney Int. 57:401-404(2000).
CC   -!- FUNCTION: Seems to play a role in the development or function of
CC       the kidney glomerular filtration barrier. May anchor the podocyte
CC       slit diaphragm to the actin cytoskeleton.
CC   -!- SUBUNIT: Interacts with podocin/NPHS2. Interacts with CD2AP C-
CC       terminal domain (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC       Predominantly located at podocyte slit diaphragm between podocyte
CC       foot processes. Also associated with podocyte apical plasma
CC       membrane.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/alpha; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Specifically expressed in podocytes of kidney
CC       glomeruli.
CC   -!- DEVELOPMENTAL STAGE: In 23-week-old embryo found in epithelial
CC       podocytes of the periphery of mature and developing glomeruli.
CC   -!- PTM: Phosphorylated on tyrosine residues.
CC   -!- DISEASE: Defects in NPHS1 are the cause of congenital nephrotic
CC       syndrome of the finnish type (NPHS1 or CNF); an autosomal
CC       recessive disorder characterized by massive proteinuria in utero
CC       and nephrosis at birth.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 IMMUNOGLOBULIN-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF035835; AAC39687.1; -.
DR   EMBL; AF190637; AAG17141.1; -.
DR   EMBL; AF126957; AAF36451.1; -.
DR   Genew; HGNC:7908; NPHS1.
DR   MIM; 602716; -.
DR   MIM; 256300; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 3.
KW   Cell adhesion; Transmembrane; Signal; Glycoprotein;
KW   Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing;
```

```
KW   Disease mutation; Polymorphism.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1241       NEPHRIN.
FT   DOMAIN       23   1055       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1056   1076       POTENTIAL.
FT   DOMAIN     1077   1241       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       46    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      153    224       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      258    324       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      354    424       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      458    535       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      560    630       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      754    823       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      856    927       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      941   1025       FIBRONECTIN TYPE-III.
FT   DOMAIN     1160   1241       BINDS TO NPHS2.
FT   DISULFID     53    111       POTENTIAL.
FT   DISULFID    160    217       POTENTIAL.
FT   DISULFID    265    317       POTENTIAL.
FT   DISULFID    361    417       POTENTIAL.
FT   DISULFID    465    528       POTENTIAL.
FT   DISULFID    567    623       POTENTIAL.
FT   DISULFID    761    816       POTENTIAL.
FT   DISULFID    863    920       POTENTIAL.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    577    577       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    908    908       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1056   1095       MISSING (IN ISOFORM 2).
FT   VARIANT      64     64       W -> S (IN CNF).
FT                                /FTId=VAR_013029.
FT   VARIANT     117    117       E -> K.
FT                                /FTId=VAR_013030.
FT   VARIANT     171    171       I -> N (IN CNF).
FT                                /FTId=VAR_013031.
FT   VARIANT     172    172       MISSING (IN CNF).
FT                                /FTId=VAR_013032.
FT   VARIANT     173    173       I -> N (IN CNF).
FT                                /FTId=VAR_013033.
FT   VARIANT     205    207       TPR -> I (IN CNF).
FT                                /FTId=VAR_013034.
FT   VARIANT     270    270       G -> C (IN CNF).
FT                                /FTId=VAR_013035.
FT   VARIANT     350    350       S -> P (IN CNF).
FT                                /FTId=VAR_013036.
FT   VARIANT     366    366       S -> R (IN CNF).
FT                                /FTId=VAR_013037.
FT   VARIANT     367    367       R -> C (IN CNF).
FT                                /FTId=VAR_013038.
FT   VARIANT     368    368       P -> S (IN CNF).
FT                                /FTId=VAR_013039.
FT   VARIANT     376    376       L -> V (IN CNF).
FT                                /FTId=VAR_013040.
FT   VARIANT     408    408       R -> Q (IN CNF; COULD BE A POLYMORPHISM).
FT                                /FTId=VAR_013041.
FT   VARIANT     447    447       E -> K (IN CNF).
FT                                /FTId=VAR_013042.
FT   VARIANT     465    465       C -> Y (IN CNF).
FT                                /FTId=VAR_013043.
FT   VARIANT     528    528       C -> F (IN CNF).
FT                                /FTId=VAR_013044.
FT   VARIANT     610    610       L -> Q (IN CNF).
FT                                /FTId=VAR_013045.
FT   VARIANT     623    623       C -> F (IN CNF).
FT                                /FTId=VAR_013046.
FT   VARIANT     724    724       S -> C (IN CNF).
FT                                /FTId=VAR_013047.
FT   VARIANT     743    743       R -> C (IN CNF).
FT                                /FTId=VAR_013048.
FT   VARIANT     802    802       R -> W (IN CNF).
FT                                /FTId=VAR_013049.
FT   VARIANT     802    802       R -> P (IN CNF).
FT                                /FTId=VAR_013050.
FT   VARIANT     806    806       A -> D (IN CNF).
FT                                /FTId=VAR_013051.
FT   VARIANT     819    819       D -> V (IN CNF).
FT                                /FTId=VAR_013052.
FT   VARIANT     831    831       R -> C (IN CNF).
FT                                /FTId=VAR_013053.

  Query Match             12.2%;  Score 76.5;  DB 1;  Length 1241;
  Best Local Similarity   28.4%;  Pred. No. 7.4;
  Matches   29;  Conservative   16;  Mismatches   34;  Indels   23;  Gaps    6;

Qy     26 SSVALNCTAWVVSGPHCSLPSV--QWLKDGLPLGI-GGHYSLHEYSWVKANLSEVLVSSV 82
Db    857 SSATLHCRA-------RGVPNIVFTWTKNGVPLDLQDPRYTEHTYHQGGVH------SSL 903

Qy     83 LGV-NVTSTEVYGAFTCSIQNISFSSFT------LQRAGPTS 117
Db    904 LTIANVSAAQDYALFTCTATNALGSDQTNIQLVSISRPDPPS 945


RESULT 12
PGH1_MOUSE
ID   PGH1_MOUSE     STANDARD;      PRT;   602 AA.
AC   P22437;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Prostaglandin G/H synthase 1 precursor (EC 1.14.99.1) (Cyclooxygenase
DE   -1) (COX-1) (Prostaglandin-endoperoxide synthase 1) (Prostaglandin H2
DE   synthase 1) (PGH synthase 1) (PGHS-1) (PHS 1).
GN   PTGS1 OR COX1 OR COX-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90203007; PubMed=2108169;
RA   Dewitt D.L., El-Harith E.A., Kraemer S.A., Andrews M.J., Yao E.F.,
RA   Armstrong R.L., Smith W.L.;
RT   "The aspirin and heme-binding sites of ovine and murine prostaglandin
RT   endoperoxide synthases.";
RL   J. Biol. Chem. 265:5192-5198(1990).
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING
CC       CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED
CC       CELLS.
CC   -!- CATALYTIC ACTIVITY: Arachidonate + AH(2) + 2 O(2) = prostaglandin
CC       H2 + A + H(2)O.
CC   -!- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND
CC       THROMBOXANES.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated. Microsomal membrane.
CC   -!- MISCELLANEOUS: THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A
CC       PEROXIDASE.
CC   -!- MISCELLANEOUS: THIS ENZYME IS THE TARGET OF NONSTEROIDAL
CC       ANTI-INFLAMMATORY DRUGS SUCH AS ASPIRIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
```

```
DR   EMBL; M34141; AAA39913.1; -.
DR   PIR; A35564; A35564.
DR   HSSP; P05979; 1DIY.
DR   MGD; MGI:97797; Ptgs1.
DR   InterPro; IPR002007; Anim_peroxidase.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF03098; An_peroxidase; 1.
DR   PRINTS; PR00457; ANPEROXIDASE.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
KW   Oxidoreductase; Dioxygenase; Peroxidase; Glycoprotein; Acetylation;
KW   Prostaglandin biosynthesis; Heme; Iron; Signal; Membrane;
KW   EGF-like domain.
FT   SIGNAL       1     26
FT   CHAIN       27    602       PROSTAGLANDIN G/H SYNTHASE 1.
FT   DOMAIN      34     72       EGF-LIKE.
FT   ACT_SITE   209    209       DISTAL HISTIDINE (BY SIMILARITY).
FT   ACT_SITE   387    387       CYCLOOXYGENASE (BY SIMILARITY).
FT   BINDING    390    390       PROXIMAL HEME LIGAND (BY SIMILARITY).
FT   MOD_RES    532    532       ASPIRIN-ACETYLATED SERINE.
FT   DISULFID    38     49       BY SIMILARITY.
FT   DISULFID    43     59       BY SIMILARITY.
FT   DISULFID    61     71       BY SIMILARITY.
FT   DISULFID    39    161       BY SIMILARITY.
FT   DISULFID   571    577       BY SIMILARITY.
FT   CARBOHYD    70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   106    106       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   146    146       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   602 AA;  69042 MW;  634C0E602045C3A0 CRC64;

  Query Match             12.1%;  Score 76;  DB 1;  Length 602;
  Best Local Similarity   27.4%;  Pred. No. 3.6;
  Matches   34;  Conservative   15;  Mismatches   59;  Indels   16;  Gaps    6;

Qy       2 PGVCDRAPDFLSPSEDQVLRPALGSSVAL---------NCTAWVVSGPHCSLPSV-QWLK 51
           | |    |   ||       |     | :         :||    |||:|::| :  ||:
Db      22 PSVLLADPGVPSPVNPCCYYPCQNQGVCVRFGLDNYQCDCTRTGYSGPNCTIPEIWTWLR 81

Qy      52 DGL-PLGIGGHYSL-HEY-SWVKAN---LSEVLVSSVLGVNVTSTEVYGAFTCSIQNISF 105
           : | |     |: | | |  |   |   : |||:  || |          :  :   ||:
Db      82 NSLRPSPSFTHFLLTHGYWLWEFVNATFIREVLMRLVLTVRSNLIPSPPTYNSAHDYISW 141

Qy     106 SSFT 109
             ||:
Db     142 ESFS 145


ENTRY 13
FGR1_HUMAN
ID   FGR1_HUMAN     STANDARD;      PRT;   822 AA.
AC   P11362; P17049;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Basic fibroblast growth factor receptor 1 precursor (EC 2.7.1.112)
DE   (FGFR-1) (bFGF-R) (Fms-like tyrosine kinase-2) (c-fgr).
GN   FGFR1 OR FLG OR FGFBR OR FLT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90245600; PubMed=2159626;
RA   Isacchi A., Bergonzoni L., Sarmientos P.;
RT   "Complete sequence of a human receptor for acidic and basic
RT   fibroblast growth factors.";
RL   Nucleic Acids Res. 18:1906-1906(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neonatal brain stem;
RX   MEDLINE=90360977; PubMed=1697263;
RA   Dionne C.A., Crumley G.R., Bellot F., Kaplow J.M., Searfoss G.,
RA   Ruta M., Burgess W.H., Jaye M., Schlessinger J.;
RT   "Cloning and expression of two distinct high-affinity receptors
RT   cross-reacting with acidic and basic fibroblast growth factors.";
RL   EMBO J. 9:2685-2692(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92282615; PubMed=1317750;
RA   Hattori Y., Odagiri H., Katoh O., Sakamoto H., Morita T.,
RA   Shimotohno K., Tobinai K., Sugimura T., Terada M.;
RT   "K-sam-related gene, N-sam, encodes fibroblast growth factor receptor
RT   and is expressed in T-lymphocytic tumors.";
RL   Cancer Res. 52:3367-3371(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91126480; PubMed=1846977;
RA   Hou J., Kan M., McKeehan K., McBride G., Adams P., McKeehan W.L.;
RT   "Fibroblast growth factor receptors from liver vary in three
RT   structural domains.";
RL   Science 251:665-668(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92118399; PubMed=1662973;
RA   Kiefer M.C., Baird A., George-Nascimento C., Nguyen T., Mason O.B.,
RA   Boley L.J., Valenzuela P., Barr P.J.;
RT   "Molecular cloning of a human basic fibroblast growth factor receptor
RT   cDNA and expression of a biologically active extracellular domain in
RT   a baculovirus system.";
RL   Growth Factors 5:115-127(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90290512; PubMed=2162671;
RA   Itoh N., Terachi T., Ohta M., Seo M.K.;
RT   "The complete amino acid sequence of the shorter form of human basic
RT   fibroblast growth factor receptor deduced from its cDNA.";
RL   Biochem. Biophys. Res. Commun. 169:680-685(1990).
RN   [7]
RP   SEQUENCE OF 201-822 FROM N.A.
RA   Ruta M., Howk R., Ricca G., Drohan W., Zabelshansky M., Laureys G.,
RA   Barton D.E., Francke U., Schlessinger J., Givol D.;
RT   "A novel protein tyrosine kinase gene whose expression is modulated
RT   during endothelial cell differentiation.";
RL   Oncogene 3:9-15(1988).
RN   [8]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=90355989; PubMed=2167437;
RA   Johnson D.E., Lee P.L., Lu J., Williams L.T.;
RT   "Diverse forms of a receptor for acidic and basic fibroblast growth
RT   factors.";
RL   Mol. Cell. Biol. 10:4728-4736(1990).
RN   [9]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=91141499; PubMed=1847500;
RA   Gutkind S.J., Link D.C., Katamine S., Lacal P., Miki T., Ley T.J.,
RA   Robbins K.C.;
RT   "A novel c-fgr exon utilized in Epstein-Barr virus-infected B
RT   lymphocytes but not in normal monocytes.";
RL   Mol. Cell. Biol. 11:1500-1507(1991).
RN   [10]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Lung;
RX   MEDLINE=91319400; PubMed=1650441;
RA   Eisemann A., Ahn J.A., Graziani G., Tronick S.R., Ron D.;
RT   "Alternative splicing generates at least five different isoforms of
RT   the human basic-FGF receptor.";
RL   Oncogene 6:1195-1202(1991).
RN   [11]
RP   SEQUENCE FROM N.A.
RA   Wennstroem S., Sandstroem C., Claesson-Welsh L.;
```

```
RL   Submitted (JUL-1990) to the EMBL/GenBank/DDBJ databases.
RN   [12]
RP   MUTAGENESIS OF TYR-766.
RX   MEDLINE=92357144; PubMed=1379697;
RA   Peters K.G., Marie J., Wilson E., Ives H.E., Escobedo J.,
RA   del Rosario M., Mirda D., Williams L.T.;
RT   "Point mutation of an FGF receptor abolishes phosphatidylinositol
RT   turnover and Ca2+ flux but not mitogenesis.";
RL   Nature 358:678-681(1992).
RN   [13]
RP   MUTAGENESIS OF TYR-766.
RX   MEDLINE=92357145; PubMed=1379698;
RA   Mohammadi M., Dionne C.A., Li W., Lin N., Spivak T., Honegger A.M.,
RA   Jaye M., Schlessinger J.;
RT   "Point mutation in FGF receptor eliminates phosphatidylinositol
RT   hydrolysis without affecting mitogenesis.";
RL   Nature 358:681-684(1992).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 464-762.
RX   MEDLINE=96361355; PubMed=8752212;
RA   Mohammadi M., Schlessinger J., Hubbard S.R.;
RT   "Structure of the FGF receptor tyrosine kinase domain reveals a novel
RT   autoinhibitory mechanism.";
RL   Cell 86:577-587(1996).
RN   [15]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 464-762.
RX   MEDLINE=97284786; PubMed=9139660;
RA   Mohammadi M., McMahon G., Sun L., Tang C., Hirth P., Yeh B.K.,
RA   Hubbard S.R., Schlessinger J.;
RT   "Structures of the tyrosine kinase domain of fibroblast growth factor
RT   receptor in complex with inhibitors.";
RL   Science 276:955-960(1997).
RN   [16]
RP   VARIANT PFEIFFER SYNDROME ARG-252.
RX   MEDLINE=95179173; PubMed=7874169;
RA   Muenke M., Schell U., Hehr A., Robin N.H., Losken H.W., Schinzel A.,
RA   Pulleyn L.J., Rutland P., Reardon W., Malcolm S., Winter R.M.;
RT   "A common mutation in the fibroblast growth factor receptor 1 gene in
RT   Pfeiffer syndrome.";
RL   Nat. Genet. 8:269-274(1994).
CC   -!- FUNCTION: RECEPTOR FOR BASIC FIBROBLAST GROWTH FACTOR. A SHORTER
CC       FORM OF THE RECEPTOR COULD BE A RECEPTOR FOR ACIDIC FGF (AFGF).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: MANY FORMS OF FGFR1 ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE FORM SHOWN HERE IS KNOWN AS ALPHA-A1.
CC   -!- DISEASE: DEFECTS IN FGFR1 ARE ONE OF THE CAUSES OF PFEIFFER
CC       SYNDROME (FS) (ALSO KNOWN AS ACROCEPHALOSYNDACTYLY TYPE V; ACS5);
CC       CHARACTERIZED BY CRANIOSYNOSTOSIS (PREMATURE FUSION OF THE SKULL
CC       SUTURES) WITH DEVIATION AND ENLARGEMENT OF THE THUMBS AND GREAT
CC       TOES, BRACHYMESOPHALANGY, WITH PHALANGEAL ANKYLOSIS AND A VARYING
CC       DEGREE OF SOFT TISSUE SYNDACTYLY.
CC   -!- DISEASE: Involved in a t(8;13)(p12;q12) chromosomal translocation
CC       which involves FGFR1 AND ZNF198. The resulting transcript is a
CC       possible candidate for stem cell leukemia lymphoma syndrome/SCLL.
CC   -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/FGFR1113.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51803; CAA36101.1; -.
DR   EMBL; X52833; CAA37015.1; -.
DR   EMBL; X66945; CAA47375.1; -.
DR   EMBL; Y00665; CAA68679.1; -.
DR   EMBL; M37722; AAA75007.1; -.
DR   EMBL; M60485; AAA35840.1; -.
DR   EMBL; M63887; AAA35958.1; -.
DR   EMBL; M34185; AAA35836.1; -.
DR   EMBL; M34186; AAA35837.1; -.
DR   EMBL; X57118; CAA40400.1; ALT_TERM.
DR   EMBL; X57119; CAA40401.1; -.
DR   EMBL; X57120; CAA40402.1; -.
DR   EMBL; X57121; CAA40403.1; -.
DR   EMBL; X57122; CAA40404.1; -.
DR   EMBL; M34641; AAA35835.1; -.
DR   EMBL; A29216; CAA01958.1; -.
DR   PIR; S11692; TVHUFG.
DR   PIR; S25420; S25420.
DR   PIR; S26739; S26739.
DR   PDB; 1FGK; 23-JUL-97.
DR   PDB; 1FGI; 08-APR-98.
DR   PDB; 1AGW; 25-MAR-98.
DR   Genew; HGNC:3688; FGFR1.
DR   MIM; 136350; -.
DR   MIM; 101600; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00219; TyrKc; 1.

  Query Match             12.1%;  Score 75.5;  DB 1;  Length 822;
  Best Local Similarity   31.3%;  Pred. No. 5.8;
  Matches   26;  Conservative    9;  Mismatches   31;  Indels   17;  Gaps    4;

Qy      23 ALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSWVKANLSEVLVSSV 82
           ||||:|   |   | | |    | :|||         |  ::       ||  | :
Db     268 ALGSNVEFMCK--VYSDPQ---PHIQWLK---------HIEVNGSKIGPDNLPYVQILKT 313

Qy      83 LGVNVTSTEVYGAFTCSIQNISF 105
              ||| |  |:       ::|:||
Db     314 AGVNTTDKEME---VLHLRNVSF 333


RESULT 14
FGR1_MOUSE
ID   FGR1_MOUSE     STANDARD;      PRT;   822 AA.
AC   P16092; Q01736; Q61562;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Basic fibroblast growth factor receptor 1 precursor (EC 2.7.1.112)
DE   (FGFR-1) (bFGF-R) (MFR).
GN   FGFR1 OR FLG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=90160373; PubMed=1689490;
RA   Reid H.H., Wilks A.F., Bernard O.;
RT   "Two forms of the basic fibroblast growth factor receptor-like mRNA
RT   are expressed in the developing mouse brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1596-1600(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=90265603; PubMed=2161096;
RA   Safran A., Avivi A., Orr-Urtereger A., Neufeld G., Lonai P.,
```

```
RA   Givol D., Yarden Y.;
RT   "The murine flg gene encodes a receptor for fibroblast growth
RT   factor.";
RL   Oncogene 5:635-643(1990).
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91207411; PubMed=1708247;
RA   Kouhara H., Kasayama S., Saito H., Matsumoto K., Sato B.;
RT   "Expression cDNA cloning of fibroblast growth factor (FGF) receptor
RT   in mouse breast cancer cells: a variant form in FGF-responsive
RT   transformed cells.";
RL   Biochem. Biophys. Res. Commun. 176:31-37(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RX   MEDLINE=90272715; PubMed=2161540;
RA   Mansukhani A., Moscatelli D., Talarico D., Levytska V., Basilico C.;
RT   "A murine fibroblast growth factor (FGF) receptor expressed in CHO
RT   cells is activated by basic FGF and Kaposi FGF.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4378-4382(1990).
RN   [5]
RP   SEQUENCE OF 1-15 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=95100926; PubMed=7802632;
RA   Harada T., Saito H., Kouhara H., Kurebayashi S., Kasayama S.,
RA   Terakawa N., Kishimoto T., Sato B.;
RT   "Murine fibroblast growth factor receptor 1 gene generates multiple
RT   messenger RNAs containing two open reading frames via alternative
RT   splicing.";
RL   Biochem. Biophys. Res. Commun. 205:1057-1063(1994).
CC   -!- FUNCTION: RECEPTOR FOR BASIC FIBROBLAST GROWTH FACTOR. A SHORTER
CC       FORM OF THE RECEPTOR COULD BE A RECEPTOR FOR ACIDIC FGF (AFGF).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A LONG FORM (SHOWN HERE), A
CC       VARIANT FORM AND A SHORT FORM ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28998; AAA37290.1; -.
DR   EMBL; X51893; CAA36175.1; -.
DR   EMBL; M65053; AAA37620.1; -.
DR   EMBL; M33760; AAA37622.1; -.
DR   EMBL; S74765; AAB32845.1; ALT_SEQ.
DR   PIR; A34849; TVMSFG.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95522; Fgfr1.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
KW   Transferase; Phosphorylation; Transmembrane; Signal;
KW   Immunoglobulin domain; Repeat; Alternative splicing.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN         22    822      BASIC FIBROBLAST GROWTH FACTOR RECEPTOR
FT                                1.
FT   DOMAIN        22    376      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      377   397      POTENTIAL.
FT   DOMAIN        398   822      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        48    108      IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        171   237      IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        270   348      IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        478   767      PROTEIN KINASE.
FT   NP_BIND       484   492      ATP (BY SIMILARITY).
FT   BINDING       514   514      ATP (BY SIMILARITY).
FT   ACT_SITE      623   623      BY SIMILARITY.
FT   MOD_RES       654   654      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   DISULFID      55    101      POTENTIAL.
FT   DISULFID      178   230      POTENTIAL.
FT   DISULFID      277   341      POTENTIAL.
FT   CARBOHYD      77    77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      117   117      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      227   227      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      240   240      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      264   264      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      296   296      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      317   317      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      330   330      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      30    30       Q -> QGSSSWPLWVAAA (IN VARIANT FORM).
FT   VARSPLIC      31    119      MISSING (IN SHORT ISOFORM).
FT   VARSPLIC      148   149      MISSING (IN VARIANT FORM).
FT   CONFLICT      229   229      T -> S (IN REF. 4).
FT   CONFLICT      256   258      ILQ -> HPS (IN REF. 1 AND 3).
FT   CONFLICT      270   270      G -> A (IN REF. 4).
FT   CONFLICT      387   387      I -> M (IN REF. 3).
FT   CONFLICT      440   440      G -> A (IN REF. 2).
FT   CONFLICT      508   508      V -> L (IN REF. 3).
FT   CONFLICT      544   544      I -> M (IN REF. 4).
FT   CONFLICT      756   756      R -> H (IN REF. 1).
FT   CONFLICT      765   765      E -> D (IN REF. 4).
SQ   SEQUENCE   822 AA;  91980 MW;  D5A4695FA680926B CRC64;

  Query Match             12.1%;  Score 75.5;  DB 1;  Length 822;
  Best Local Similarity   31.3%;  Pred. No. 5.8;
  Matches   26;  Conservative    9;  Mismatches   31;  Indels   17;  Gaps    4;

Qy      23 ALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSWVKANLSEVLVSSV 82
           ||||:|   |   | | |   | :||||         |  ::       ||  | :
Db     268 ALGSNVEFMCK--VYSDPQ---PHIQWLK---------HIEVNGSKIGPDNLPYVQILKT 313

Qy      83 LGVNVTSTEVYGAFTCSIQNISF 105
             ||| |  |:       ::|:||
Db     314 AGVNTTDKEME---VLHLRNVSF 333


RESULT 15
LACH_DROME
ID   LACH_DROME     STANDARD;      PRT;   359 AA.
AC   Q24372;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lachesin precursor.
GN   LAC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94038693; PubMed=8223276;
RA   Karlstrom R.O., Wilder L.P., Bastiani M.J.;
RT   "Lachesin: an immunoglobulin superfamily protein whose expression
RT   correlates with neurogenesis in grasshopper embryos.";
RL   Development 118:509-522(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND
```

```
CC       AXON OUTGROWTH.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS
CC       FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEMS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13255; AAC37184.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0010238; Lac.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    336       LACHESIN.
FT   PROPEP      337    359       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       43    120       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      150    211       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      240    310       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     50    113       POTENTIAL.
FT   DISULFID    157    204       POTENTIAL.
FT   DISULFID    247    303       POTENTIAL.
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       336    336       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   359 AA;  39998 MW;  D1F513E2B9D886E6 CRC64;

  Query Match             12.0%;  Score 75;  DB 1;  Length 359;
  Best Local Similarity   25.3%;  Pred. No. 2.6;
  Matches   23;  Conservative   11;  Mismatches   37;  Indels   20;  Gaps    4;

       19 VLRPALGSSVA----LNCTAWVVSGPHCSL---PSVQWLKDGLPLGIGGHYSLHEYSWVK 71
          | || || ::     | |         |       |:: | || : |    |||:  ::
Db    230 VPRPRLGQALQYDMDLEC--------HIEAYPPPAIVWTKDDIQLANNQHYSISHFA--- 278

Qy     72 ANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
            ::    | | |       || : |   |
Db    279 --TADEYTDSTLRVITVEKRQYGDYVCKATN 307


Search completed: November 14, 2002, 17:30:20
Job time : 16 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 14, 2002, 17:27:23 ; Search time 33 Seconds
(without alignments)
736.775 Million cell updates/sec

Title: US-09-598-443-2_COPY_1_118
Perfect score: 626
Sequence: 1 MPGVCDRAPDFLSPSEDQVL.....SIQNISFSSFTLQRAGPTSH 118

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	626	100.0	410	4	Q9H733	Q9h733 homo sapien
2	407.5	65.1	409	11	Q9JLZ8	Q9jlz8 mus musculu
3	105	16.8	1319	4	Q9HCD3	Q9hcd3 homo sapien
4	96	15.3	534	5	Q9VSG5	Q9vsg5 drosophila
5	94.5	15.1	295	5	Q9VS29	Q9vs29 drosophila
6	94	15.0	793	11	O70246	O70246 mus musculu
7	93.5	14.9	7962	4	Q10465	Q10465 homo sapien
8	93.5	14.9	34350	4	Q8WZ42	Q8wz42 homo sapien
9	91.5	14.6	1323	13	Q08476	Q08476 gallus gall
10	90	14.4	6620	4	Q96AA2	Q96aa2 homo sapien
11	90	14.4	6831	5	Q23550	Q23550 caenorhabdi
12	90	14.4	7160	5	Q23551	Q23551 caenorhabdi
13	88	14.1	1036	5	Q8SWW3	Q8sww3 drosophila
14	88	14.1	1336	5	Q9VN14	Q9vn14 drosophila
15	88	14.1	2673	4	Q96SC3	Q96sc3 homo sapien
16	88	14.1	5636	4	Q96RW7	Q96rw7 homo sapien
17	87.5	14.0	5198	5	O76518	O76518 caenorhabdi
18	84	13.4	298	4	Q96HT1	Q96ht1 homo sapien
19	84	13.4	836	4	O94856	O94856 homo sapien
20	84	13.4	1174	11	Q91Z60	Q91z60 rattus norv
21	84	13.4	2053	4	Q8WXU7	Q8wxu7 homo sapien
22	84	13.4	2113	4	Q8TD84	Q8td84 homo sapien
23	82.5	13.2	500	5	Q9XZB7	Q9xzb7 drosophila
24	82.5	13.2	4824	5	Q95YM1	Q95ym1 procambarus
25	82.5	13.2	17352	5	Q95YM2	Q95ym2 procambarus
26	82	13.1	582	11	Q8R4B5	Q8r4b5 mus musculu
27	82	13.1	915	11	Q8R4B3	Q8r4b3 mus musculu
28	81.5	13.0	496	5	Q9W260	Q9w260 drosophila
29	81.5	13.0	500	5	Q961W0	Q961w0 drosophila
30	81.5	13.0	650	11	Q63709	Q63709 rattus ratt
31	81.5	13.0	1419	13	Q98SW3	Q98sw3 brachydanio
32	81	12.9	1151	11	Q9QVN5	Q9qvn5 rattus sp.
33	81	12.9	1217	11	P97685	P97685 rattus norv
34	80.5	12.9	6875	6	Q28733	Q28733 oryctolagus
35	80	12.8	1746	4	Q8WY19	Q8wy19 homo sapien
36	80	12.8	4650	4	Q15598	Q15598 homo sapien
37	80	12.8	6632	5	Q17362	Q17362 caenorhabdi
38	80	12.8	6632	5	O01761	O01761 caenorhabdi
39	80	12.8	26926	4	Q10466	Q10466 homo sapien
40	80	12.8	26926	4	Q8WZB3	Q8wzb3 homo sapien
41	79.5	12.7	137	13	Q9YHS0	Q9yhs0 ginglymosto
42	79.5	12.7	868	11	Q62838	Q62838 rattus norv
43	79.5	12.7	1256	11	Q9ET59	Q9et59 mus musculu
44	79.5	12.7	1256	11	Q9JIX1	Q9jix1 mus musculu
45	79.5	12.7	1256	11	Q925S5	Q925s5 mus musculu

ALIGNMENTS

RESULT 1
Q9H733
ID Q9H733 PRELIMINARY; PRT; 410 AA.
AC Q9H733;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA: FLJ21446 fis, clone COL04458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK025099; BAB15066.1; -.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR000157; TIR_domain.
DR Pfam; PF01582; TIR; 1.
DR PRINTS; PR01559; DUFFYANTIGEN.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00255; TIR; 1.
SQ SEQUENCE 410 AA; 45707 MW; 2A7A663D79567ED6 CRC64;

Query Match 100.0%; Score 626; DB 4; Length 410;
Best Local Similarity 100.0%; Pred. No. 6e-60;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60

Qy      61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118


RESULT 2
Q9JLZ8
ID   Q9JLZ8       PRELIMINARY;      PRT;   409 AA.
AC   Q9JLZ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Toll/interleukin-1 receptor 8.
GN   SIGIRR OR TIR8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Penton-Rol G., Polentarutti N., Muzio M., Mantovani A.;
RT   "TIR8: a novel toll/interleukin-1 receptor family member.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF113795; AAF26200.1; -.
DR   MGD; MGI:1344402; Sigirr.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000157; TIR_domain.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF01582; TIR; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00255; TIR; 1.
KW   Receptor.
SQ   SEQUENCE   409 AA;  45707 MW;  DC4AB99A72356A13 CRC64;

  Query Match             65.1%;  Score 407.5;  DB 11;  Length 409;
  Best Local Similarity   67.8%;  Pred. No. 4.1e-36;
  Matches   80;  Conservative    8;  Mismatches   29;  Indels    1;  Gaps    1;

Qy       1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGG 60
           | |||| ||:||||||||| |  |||  |||||||||| | | |  ||||||||||| || |
Db       1 MAGVCDMAPNFLSPSEDQALGLALGREVALNCTAWVFSRPQCPQPSVQWLKDGLALGNGS 60

Qy      61 HYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGPTSH 118
           |:|||  || || ||: ||||| :|:|: | || ||||: |:|  |||| ||||  |
Db      61 HFSLHEDFWVSANFSEI-VSSVLVLNLTNAEDYGTFTCSVWNVSSHSFTLWRAGPAGH 117


RESULT 3
Q9HCD3
ID   Q9HCD3       PRELIMINARY;      PRT;  1319 AA.
AC   Q9HCD3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   KIAA1639 protein (Fragment).
GN   KIAA1639.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20450683; PubMed=10997877;
RA   Nagase T., Kikuno R., Nakayama M., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.
RT   XVIII. The complete sequences of 100 new cDNA clones from brain which
RT   code for large proteins in vitro.";
RL   DNA Res. 7:273-281(2000).
DR   EMBL; AB046859; BAB13465.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00069; pkinase; 2.
DR   ProDom; PD000001; Euk_pkinase; 2.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
KW   Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   1319 AA;  142049 MW;  FBA43AE17204EF48 CRC64;

  Query Match             16.8%;  Score 105;  DB 4;  Length 1319;
  Best Local Similarity   31.4%;  Pred. No. 0.017;
  Matches   33;  Conservative   11;  Mismatches   29;  Indels   32;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQ--WLKDGLPLGIGGHYS 63
           |||| ||    |: :   || || | |          : |: |  | ||| ||
Db     813 DRAPTFLRELSDETV--VLGQSVTLACQV-------SAQPAAQATWSKDGAPL------- 856

Qy      64 LHEYSWVKANLSEVLVSS------VLGVNVTSTEVYGAFTCSIQN 102
                 : | ||:|:      :| : |   |  | :|||: |
Db     857 --------ESSSRVLISATLKNFQLLTILVVVAEDLGVYTCSVSN 893


RESULT 4
Q9VSG5
ID   Q9VSG5       PRELIMINARY;      PRT;   534 AA.
AC   Q9VSG5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG18630 protein.
GN   CG18630.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003556; AAF50456.1; -.
DR   FlyBase; FBgn0035863; CG18630.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 4.
KW   Immunoglobulin domain.
SQ   SEQUENCE   534 AA;  59845 MW;  E22071BAE89D2F66 CRC64;

  Query Match             15.3%;  Score 96;  DB 5;  Length 534;
  Best Local Similarity   29.6%;  Pred. No. 0.055;
  Matches   29;  Conservative   17;  Mismatches   38;  Indels   14;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLH 65
           | :|: |    :| |:|  | :|:| | |     |   ||   |  || |:    : :
Db     156 DASPELLYWFSEQTLQP--GPTVSLKCVA--TGNP---LPQFTWSLDGFPIPDSSRFLVG 208

Qy      66 EYSWVKAN-LSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
           :|  :  : :| | :|:|        |  | :||: ||
Db     209 QYVTIHDDVISHVNISNV------KEEDGGEYTCTAQN 240


RESULT 5
Q9VS29
ID   Q9VS29      PRELIMINARY;      PRT;   295 AA.
AC   Q9VS29;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG8618 protein.
GN   CG8618.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003560; AAF50601.1; -.
DR   FlyBase; FBgn0035748; CG8618.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 4.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   295 AA;  32290 MW;  48F455CE162C4FE1 CRC64;

  Query Match             15.1%;  Score 94.5;  DB 5;  Length 295;
  Best Local Similarity   28.2%;  Pred. No. 0.038;
  Matches   29;  Conservative   20;  Mismatches   39;  Indels   15;  Gaps    5;

Qy       1 MPGVCDRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLP-SVQWLKDGLPLGIG 59
             :|    : |  ||| :  :|:  :|   :| |:  || |    || :: | ||| |:
Db      91 LPDDIRQLPPKLSPFQTNILQLNMGDRASLTCS--VVKG---DLPLTINWRKDGRPIDPT 145

Qy      60 GHYSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
           | |: :       : :||   || : |     | ::| ::|
Db     146 QHMSVKQVD----QYNSILVIENLGSDHT-----GNYSCVVRN 179


RESULT 6
O70246
ID   O70246      PRELIMINARY;      PRT;   793 AA.
AC   O70246;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Putative neuronal cell adhesion molecule (PUNC) (Putative neuronal
DE   cell adhesion molecule, short form).
GN   PUNC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RX   MEDLINE=98175891; PubMed=9507132;
RA   Salbaum J.M.;
RT   "Punc, a novel mouse gene of the immunoglobulin superfamily, is
RT   expressed predominantly in the developing nervous system.";
RL   Mech. Dev. 71:201-204(1998).
RN   [2]
RP   SEQUENCE OF 599-715 FROM N.A.
RC   STRAIN=FVB;
RX   MEDLINE=99122837; PubMed=9922388;
RA   Salbaum J.M.;
RT   "Genomic structure and chromosomal localization of the mouse gene
RT   Punc.";
RL   Mamm. Genome 10:107-111(1999).
DR   EMBL; AF026465; AAD12133.1; -.
```

```
DR   EMBL; AF026466; AAD12124.1; -.
DR   HSSP; P80362; 1WTL.
DR   MGD; MGI:1202390; Punc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 4.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 4.
KW   Immunoglobulin domain.
SQ   SEQUENCE   793 AA;  84287 MW;  65FD4D1D5E952937 CRC64;

  Query Match             15.0%;  Score 94;  DB 11;  Length 793;
  Best Local Similarity   30.7%;  Pred. No. 0.15;
  Matches   31;  Conservative   11;  Mismatches   41;  Indels   18;  Gaps    4;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           :||          : ||| |::    | |     ||     | |||:|  || |||  |
Db     339 QAPAEFVQHPQSISRPA-GTTAMFTCQAQGEPPPH-----VTWLKNGQVLGAGGHVRL-- 390

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSS 107
             | |      :| | ::    |     : |  :||: ||
Db     391 ----KNN------NSTLSISGVGPEDEAIYQCVAENIAGSS 421


RESULT 7
Q10465
ID   Q10465      PRELIMINARY;      PRT;  7962 AA.
AC   Q10465;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Titin, skeletal muscle isoform (EC 2.7.1.-) (Connectin) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RX   MEDLINE=96026330; PubMed=7569978;
RA   Labeit S., Kolmer B.;
RT   "Titins: giant proteins in charge of muscle ultrastructure and
RT   elasticity.";
RL   Science 270:293-296(1995).
CC   -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC       ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC       SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
CC       PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
CC       DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
CC       ONE TISSUE.
CC   -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC   -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC       KINASES.
CC   -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 90
CC       IMMUNOGLOBULIN C2-LIKE DOMAINS.
DR   EMBL; X90569; CAA62189.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR004168; PPAK_motif.
DR   Pfam; PF00047; ig; 59.
DR   Pfam; PF02818; PPAK; 53.
DR   SMART; SM00408; IGc2; 43.
DR   SMART; SM00410; IG_like; 15.
KW   Muscle protein; Cytoskeleton; Structural protein; Phosphorylation;
KW   Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW   Immunoglobulin domain.
FT   NON_TER       1      1
FT   DOMAIN     5618   7792       GLU/LYS/PRO/VAL-RICH.
FT   NON_TER    7962   7962
SQ   SEQUENCE   7962 AA;  883018 MW;  B85240533CBADE58 CRC64;

  Query Match             14.9%;  Score 93.5;  DB 4;  Length 7962;
  Best Local Similarity   32.8%;  Pred. No. 2.8;
  Matches   38;  Conservative    7;  Mismatches   32;  Indels   39;  Gaps    7;

Qy      10 DFLSPSEDQVLR-----------------PALGSSVALNCTAWVVSGPHCSLP-SVQWL 50
           || | | |  ||                    ||||: : |    |||   ||| | ||
Db    1883 DFGSSSCDAYLRVLDQNIPPSFTKKLTKMDKVLGSSIHMECK---VSG---SLPISAQWF 1936

Qy      51 KDGLPLGIGGHYSL--HEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNIS 104
           |||  :    | |  || |          ||  | ||    |    :|| : |::
Db    1937 KDGKEISTSAKYRLVCHERS----------VS--LEVNNLELEDTANYTCKVSNVA 1980


RESULT 8
Q8WZ42
ID   Q8WZ42      PRELIMINARY;      PRT; 34350 AA.
AC   Q8WZ42;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Titin.
GN   TTN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20309627; PubMed=10850961;
RA   Freiburg A., Trombitas K., Hell W., Cazorla O., Fougerousse F.,
RA   Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA   Granzier H., Labeit S.;
RT   "Series of exon-skipping events in the elastic spring region of titin
RT   as the structural basis for myofibrillar elastic diversity.";
RL   Circ. Res. 86:1114-1121(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21573839; PubMed=11717165;
RA   Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
RA   McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
RA   Labeit S.;
RT   "The complete gene sequence of titin, expression of an unusual ~700
RT   kDa titin isoform and its interaction with obscurin identify a novel
RT   Z-line to I-band linking system.";
RL   Circ. Res. 89:1065-1072(2001).
DR   EMBL; AJ277892; CAD12456.1; -.
DR   InterPro; IPR000282; Cytok_receptor_2.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000577; FGGY_kin.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR002016; Peroxidase.
DR   InterPro; IPR004168; PPAK_motif.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 132.
DR   Pfam; PF00047; ig; 146.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF02818; PPAK; 53.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 133.
DR   SMART; SM00409; IG; 167.
DR   SMART; SM00408; IGc2; 148.
DR   SMART; SM00406; IGv; 23.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
```

```
DR   PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
SQ   SEQUENCE   34350 AA;  3816262 MW;  5B1120058A7CE58A CRC64;

  Query Match             14.9%;  Score 93.5;  DB 4;  Length 34350;
  Best Local Similarity   32.8%;  Pred. No. 17;
  Matches   38;  Conservative    7;  Mismatches   32;  Indels   39;  Gaps    7;

Qy      10 DFLSPSEDQVLR-----------------PALGSSVALNCTAWVVSGPHCSLP-SVQWL 50
           || | | |  ||                   ||||: : |    |||   ||| | ||
Db    6145 DFGSSSCDAYLRVLDQNIPPSFTKKLTKMDKVLGSSIHMECK---VSG---SLPISAQWF 6198

Qy      51 KDGLPLGIGGHYSL--HEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNIS 104
           |||  :    | |  || |          ||  | ||    |    :|| : |::
Db    6199 KDGKEISTSAKYRLVCHERS----------VS--LEVNNLELEDTANYTCKVSNVA 6242


RESULT 9
Q08476
ID   Q08476      PRELIMINARY;      PRT;  1323 AA.
AC   Q08476;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Connectin(titin) (Fragment).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RA   Maruyama K., Endo T., Kume H., Kawamura Y., Kanzawa N., Nakauchi Y.,
RA   Kimura S., Kawashima S., Maruyama K.;
RT   "A novel domain sequence of connectin localized at the I band of
RT   skeletal muscle sarcomeres: homology ro neurofilament subunits.";
RL   Cell. Mol. Biol. Res. 194:1288-1291(1993).
DR   EMBL; D16541; BAA03979.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 11.
DR   SMART; SM00408; IGc2; 9.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
FT   NON_TER       1      1
SQ   SEQUENCE   1323 AA;  148581 MW;  3974F64082CA926B CRC64;

  Query Match             14.6%;  Score 91.5;  DB 13;  Length 1323;
  Best Local Similarity   24.5%;  Pred. No. 0.51;
  Matches   27;  Conservative   21;  Mismatches   31;  Indels   31;  Gaps    5;

Qy      19 VLRPAL----------GSSVALNCTAWVVSGPHCSLPSV--QWLKDGLPLGIGGHYSLH 65
           |: ||:          |::  |:|        |:|:   || ::| |: ||| ||
Db     959 VIEPAVERHLHDTTFKEGNTCTLSC--------QFSIPNAKSQWYRNGRPIKIGGRYS-- 1008

Qy      66 EYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQNISFSSFTLQRAGP 115
             : |   : ::::  |       ||  | :|| : |:  ::     | |
Db    1009 --TQVSDKVHKLIIKDV------RTEDQGQYTCKLDNLETTADLTIEAEP 1050


RESULT 10
Q96AA2
ID   Q96AA2      PRELIMINARY;      PRT;  6620 AA.
AC   Q96AA2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Obscurin.
GN   OBSCN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE=21342081; PubMed=11448995;
RA   Young P.W., Ehler E., Gautel M.;
RT   "Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor
RT   protein involved in sarcomere assembly.";
RL   J. Cell Biol. 154:123-136(2001).
DR   EMBL; AJ002535; CAC44768.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 49.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
SQ   SEQUENCE   6620 AA;  721665 MW;  C2AE8EB77B284452 CRC64;

  Query Match             14.4%;  Score 90;  DB 4;  Length 6620;
  Best Local Similarity   27.8%;  Pred. No. 5.4;
  Matches   27;  Conservative   24;  Mismatches   30;  Indels   16;  Gaps    5;

Qy       6 DRAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLH 65
           | || ||:  ::| ::   |  |: :|   ||:|    :|||:| |||  |    || ::
Db    5123 DAAPVFLTELQNQEVQD--GYPVSFDC---VVTGQ--PMPSVRWFKDGKLLEEDDHYMIN 5175

Qy      66 EYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
           |    :    ::::::|:  ::      | : |  :|
Db    5176 E---DQQGGHQLIITAVVPADM------GVYRCLAEN 5203


RESULT 11
Q23550
ID   Q23550      PRELIMINARY;      PRT;  6831 AA.
AC   Q23550; Q23020; Q27232;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UNC-22 protein.
GN   UNC-22 OR ZK617.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   White S., Harris B.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 784-6831 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=90044042; PubMed=2812002;
RA   Benian G.M., Kiff J.E., Neckelmann N., Moerman D.G., Waterson R.H.;
RT   "Sequence of an unusually large protein implicated in regulation of
RT   myosin activity in C. elegans.";
RL   Nature 342:45-50(1989).
RN   [3]
RP   SEQUENCE OF 784-6831 FROM N.A.
RC   STRAIN=BRISTOL N2;
```

```
RX   MEDLINE=93387664; PubMed=8397135;
RA   Benian G.M., L'Hernault S.W., Morris M.E.;
RT   "Additional sequence complexity in the muscle gene, unc-22, and its
RT   encoded protein, twitchin, of Caenorhabditis elegans.";
RL   Genetics 134:1097-1104(1993).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM ZK617.1A (SHOWN HERE) AND
CC       ISOFORM ZK617.1B (Q23551); MAY BE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; Z73897; CAA98064.1; -.
DR   EMBL; Z73899; CAA98064.1; JOINED.
DR   EMBL; Z73899; CAA98081.1; -.
DR   EMBL; Z73897; CAA98081.1; JOINED.
DR   EMBL; X15423; CAA33463.1; -.
DR   HSSP; Q63450; 1A06.
DR   WormPep; ZK617.1A; CE06634.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00041; fn3; 31.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 27.
DR   SMART; SM00408; IGc2; 6.
DR   SMART; SM00410; IG_like; 17.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Alternative splicing; Hypothetical protein;
KW   Immunoglobulin domain; Repeat; Serine/threonine-protein kinase;
KW   Transferase.
SQ   SEQUENCE   6831 AA;  752596 MW;  6CADE032963A52E1 CRC64;

  Query Match             14.4%;  Score 90;  DB 5;  Length 6831;
  Best Local Similarity   24.0%;  Pred. No. 5.6;
  Matches   23;  Conservative   21;  Mismatches   40;  Indels   12;  Gaps    3;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           | | |:   : :::    |: : : |          | |  :|:|||:|| :|| |
Db     101 RGPSFV--GKPRIIPKDGGALIVMECKV-----KSASTPVAKWMKDGVPLSMGGLYHA-- 151

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
             : ::| :      | :   |:   | : |:|:|
Db     152 ---IFSDLGDQTYLCQLEIRGPSSSDAGQYRCNIRN 184


RESULT 12
Q23551
ID   Q23551      PRELIMINARY;      PRT;  7160 AA.
AC   Q23551;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   UNC-22 protein.
GN   UNC-22 OR ZK617.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., White S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM ZK617.1A (Q23550) AND
CC       ISOFORM ZK617.1B (SHOWN HERE); MAY BE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; Z73899; CAA98082.1; -.
DR   EMBL; Z73897; CAA98082.1; JOINED.
DR   EMBL; Z73897; CAA98065.1; -.
DR   EMBL; Z73899; CAA98065.1; JOINED.
DR   HSSP; Q63450; 1A06.
DR   WormPep; ZK617.1B; CE06635.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00041; fn3; 31.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 27.
DR   SMART; SM00408; IGc2; 6.
DR   SMART; SM00410; IG_like; 18.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Alternative splicing; Hypothetical protein;
KW   Immunoglobulin domain; Repeat; Serine/threonine-protein kinase;
KW   Transferase.
SQ   SEQUENCE   7160 AA;  789230 MW;  3B7D0A4F627CD027 CRC64;

  Query Match             14.4%;  Score 90;  DB 5;  Length 7160;
  Best Local Similarity   24.0%;  Pred. No. 6;
  Matches   23;  Conservative   21;  Mismatches   40;  Indels   12;  Gaps    3;

Qy       7 RAPDFLSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHE 66
           | | |:   : :::    |: : : |          | |  :|:|||:|| :|| |
Db     101 RGPSFV--GKPRIIPKDGGALIVMECKV-----KSASTPVAKWMKDGVPLSMGGLYHA-- 151

Qy      67 YSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
             : ::| :      | :   |:   | : |:|:|
Db     152 ---IFSDLGDQTYLCQLEIRGPSSSDAGQYRCNIRN 184


RESULT 13
Q8SWW3
ID   Q8SWW3      PRELIMINARY;      PRT;  1036 AA.
AC   Q8SWW3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   LD28757p.
GN   CG1084.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY095040; AAM11368.1; -.
```

```
SQ   SEQUENCE   1036 AA;  117742 MW;  A370F297D39D719D CRC64;

  Query Match             14.1%;  Score 88;  DB 5;  Length 1036;
  Best Local Similarity   27.7%;  Pred. No. 0.92;
  Matches   28;  Conservative    9;  Mismatches   42;  Indels   22;  Gaps    4;

Qy       4 VCDRAPDFLS-PSEDQVLRPALGSSVALNCTAWVVSGPHCS-LPSVQWLKDGLPLGIGGH 61
           |    | |   | | :|          |    :|  |  :  |  || |||  :| |||
Db     397 VLSMKPSFKKHPLESEVY-------AVYNGNTTIVCDPEAAPRPKFQWKKDGQVIGSGGH 449

Qy      62 YSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
                     :| |  | :: || :  | :||   |
Db     450 ------------RRILPSGTLTISPTSRDDEGIYTCIASN 477


RESULT 14
Q9VN14
ID   Q9VN14      PRELIMINARY;      PRT;  1336 AA.
AC   Q9VN14;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG1084 protein.
GN   CG1084.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003606; AAF52137.1; -.
DR   HSSP; P20241; 1CFB.
DR   FlyBase; FBgn0037240; CG1084.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00041; fn3; 4.
DR   Pfam; PF00047; ig; 5.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00060; FN3; 4.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1336 AA;  151994 MW;  1636094A64484262 CRC64;

  Query Match             14.1%;  Score 88;  DB 5;  Length 1336;
  Best Local Similarity   27.7%;  Pred. No. 1.3;
  Matches   28;  Conservative    9;  Mismatches   42;  Indels   22;  Gaps    4;

Qy       4 VCDRAPDFLS-PSEDQVLRPALGSSVALNCTAWVVSGPHCS-LPSVQWLKDGLPLGIGGH 61
           |    | |   | | :|          |    :|  |  :  |  || |||  :| |||
Db     697 VLSMKPSFKKHPLESEVY-------AVYNGNTTIVCDPEAAPRPKFQWKKDGQVIGSGGH 749

Qy      62 YSLHEYSWVKANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
                     :| |  | :: || :  | :||   |
Db     750 ------------RRILPSGTLTISPTSRDDEGIYTCIASN 777


RESULT 15
Q96SC3
ID   Q96SC3      PRELIMINARY;      PRT;  2673 AA.
AC   Q96SC3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Fibulin-6 (Fragment).
GN   FIBL-6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Kostka G., Timpl R.;
RT   "Partial sequence of fibulin-6 with a c-terminal region related to
RT   domain II and III of the fibulin family.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ306906; CAC37630.1; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000875; Cecropin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00047; ig; 17.
DR   Pfam; PF00090; tsp_1; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_5.
DR   PROSITE; PS00268; CECROPIN; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_3.
DR   PROSITE; PS01187; EGF_CA; UNKNOWN_8.
DR   PROSITE; PS50092; TSP1; 6.
FT   NON_TER       1       1
SQ   SEQUENCE   2673 AA;  291017 MW;  BEAEC30B8340E272 CRC64;

  Query Match             14.1%;  Score 88;  DB 4;  Length 2673;
  Best Local Similarity   28.6%;  Pred. No. 2.9;
  Matches   26;  Conservative   13;  Mismatches   32;  Indels   20;  Gaps    4;

Qy      12 LSPSEDQVLRPALGSSVALNCTAWVVSGPHCSLPSVQWLKDGLPLGIGGHYSLHEYSWVK 71
```

```
          :   |  ||:   ||| :: |   :  |     ||: ||:|| |||:  | ::  :
Db    478 MGTEEITVLK---GSSTSMAC---ITDG--TPAPSMAWLRDGQPLGLDAHLTVSTH---- 525

Qy     72 ANLSEVLVSSVLGVNVTSTEVYGAFTCSIQN 102
                 || :    ||  | :||   |
Db    526 --------GMVLQLLKAETEDSGKYTCIASN 548

Search completed: November 14, 2002, 17:31:05
Job time : 42 secs
```